J Pathol Inform

Editor-in-Chief:
Anil V. Parwani , Pittsburgh, PA, USA
Liron Pantanowitz, Pittsburgh, PA, USA

OPEN ACCESS
HTML format

For entire Editorial Board visit : www.jpathinformatics.org/editorialboard.asp

Original Article

Content-based image retrieval of digitized histopathology in boosted spectrally embedded spaces

Akshay Sridhar[1], Scott Doyle[1], Anant Madabhushi[1]

[1]Department of Biomedical Engineering, Rutgers University, The State University of New Jersey, Piscataway, NJ 08854, USA

E-mail: sridhaak@rowan.edu
*Corresponding author

Received: 17 September 2013 Accepted: 04 November 2014 Published: 29 June 2015

This article may be cited as:
Sridhar A, Doyle S, Madabhushi A. Content-based image retrieval of digitized histopathology in boosted spectrally embedded spaces. J Pathol Inform 2015;6:41.

Available FREE in open access from: http://www.jpathinformatics.org/text.asp?2015/6/1/41/159441

Abstract

Context: Content-based image retrieval (CBIR) systems allow for retrieval of images from within a database that are similar in visual content to a query image. This is useful for digital pathology, where text-based descriptors alone might be inadequate to accurately describe image content. By representing images via a set of quantitative image descriptors, the similarity between a query image with respect to archived, annotated images in a database can be computed and the most similar images retrieved. Recently, non-linear dimensionality reduction methods have become popular for embedding high-dimensional data into a reduced-dimensional space while preserving local object adjacencies, thereby allowing for object similarity to be determined more accurately in the reduced-dimensional space. However, most dimensionality reduction methods implicitly assume, in computing the reduced-dimensional representation, that all features are equally important. **Aims**: In this paper we present boosted spectral embedding (BoSE), which utilizes a boosted distance metric to selectively weight individual features (based on training data) to subsequently map the data into a reduced-dimensional space. **Settings and Design**: BoSE is evaluated against spectral embedding (SE) (which employs equal feature weighting) in the context of CBIR of digitized prostate and breast cancer histopathology images. **Materials and Methods**: The following datasets, which were comprised of a total of 154 hematoxylin and eosin stained histopathology images, were used: (1) Prostate cancer histopathology (benign vs. malignant), (2) estrogen receptor (ER) + breast cancer histopathology (low vs. high grade), and (3) HER2+ breast cancer histopathology (low vs. high levels of lymphocytic infiltration). **Statistical Analysis Used**: We plotted and calculated the area under precision-recall curves (AUPRC) and calculated classification accuracy using the Random Forest classifier. **Results**: BoSE outperformed SE both in terms of CBIR-based (area under the precision-recall curve) and classifier-based (classification accuracy) on average across all of the dimensions tested for all three datasets: (1) Prostate cancer histopathology (AUPRC: BoSE = 0.79, SE = 0.63; Accuracy: BoSE = 0.93, SE = 0.80), (2) ER + breast cancer histopathology (AUPRC: BoSE = 0.79, SE = 0.68; Accuracy: BoSE = 0.96, SE = 0.96), and (3) HER2+ breast cancer histopathology (AUPRC: BoSE = 0.54, SE = 0.44; Accuracy: BoSE = 0.93, SE = 0.91). **Conclusion**: Our results suggest that BoSE could serve as an important tool for CBIR and classification of high-dimensional biomedical data.

Key words: Boosted, breast cancer, content-based image retrieval, histopathology, non-linear dimensionality reduction, prostate cancer, spectral embedding

Access this article online

Website: www.jpathinformatics.org

DOI: 10.4103/2153-3539.159441

Quick Response Code:

INTRODUCTION

Content-based image retrieval (CBIR) systems allow a user to retrieve images from a database based on visual similarity to the query image. This is particularly useful for digital pathology and medical imaging databases, where text-based descriptors alone might be inadequate to accurately describe image content.[1-7] In CBIR systems, a query image is used as the input and based on image attribute matching; the most similar images from within a database are retrieved. Two main components of a CBIR system are (a) the image (or feature) representation, and (b) choice of similarity metric for performing retrieval. An ideal similarity metric (distance measure) would yield a large value when comparing visually dissimilar images and a small value when similar images are compared. For any given query image, the most similar images in the database as determined by the similarity metric are retrieved in decreasing order of relevance. However, in cases where images are represented by a large number of image attributes, the similarity measure might be affected by the so called "curse of dimensionality" problem, wherein the number of attributes may be greater than the total number of instances in the database.

Dimensionality reduction (DR) is a technique that is used to project high-dimensional data into a reduced-dimensional embedding space. The low-dimensional data representation allows for more consistent and accurate similarity computations, compared to the high-dimensional space, to help determine image similarity[8,9] DR techniques can be broadly categorized as linear or nonlinear. Linear DR techniques such as principal component analysis (PCA)[10] fail to accurately capture object (image) relationships where the data reside on some non-linear manifold.[11] Objects residing on different ends of the manifold could potentially be mapped closer to each other in the lower-dimensional space, since linear DR methods use the Euclidean norm as opposed to the geodesic distance (appropriate for adjacency determination for objects residing on nonlinear manifolds). Nonlinear dimensionality reduction (NLDR) methods[12-15] attempt to capture object adjacency on nonlinear manifolds by preservation of the local linear neighborhood structure.[16] However, NLDR methods such as Isomap[12] and locally linear embedding (LLE)[13] are sensitive to the choice of the size of the local neighborhood (κ) within which linearity is assumed. Diffusion Maps,[14] another NLDR method, is sensitive to the number of time steps specified for the random walk. spectral embedding (SE)[15] is a NLDR method that unlike neighborhood preserving NLDR schemes (such as LLE, Isomap), defines object adjacency by using a Gaussian kernel in conjunction with the Euclidean distance metric (EDM) to yield a similarity matrix for all objects. The eigenvalue decomposition of this similarity matrix is then determined to yield the low-dimensional representation (eigenvectors) of the data. While SE is still sensitive to the parameters of the kernel, it has been shown to be more robust compared to LLE and Isomap.[17] CBIR could be performed in conjunction with SE by mapping the query and database images into a reduced-dimensional space and then retrieving relevant images as those in the neighborhood of the query instance. A key shortcoming of the EDM, however, is that it implicitly assumes all features (dimensions) are equally relevant. In the context of CBIR, features that are poor in discriminating between two image classes could potentially map dissimilar images close to each other in the low-dimensional space. Hence, in order to determine a more optimal low-dimensional representation of the data, it is desirable to weight the discriminatory attributes higher compared to the erroneous or noisy features prior to computing the similarity matrix.

There has been some previous work in the development of SE variants. Tiwari *et al*. proposed a weighted multi-kernel learning scheme to yield an improved weight matrix for use in conjunction with SE.[18] ElGhawalby and Hancock[19] formulated a variant of SE that used an edge-based wave kernel that embedded the nodes of a graph as points on the surface of a manifold, and used the resulting point-set to compute graph characteristics. Robles-Kelly and Hancock[20] used the Kruskal coordinates to compute the edge-weights for a weight matrix and used it to embed the nodes of the graph onto a Riemannian manifold.

In this paper we employ a novel variant of SE called boosted spectral embedding (BoSE), a supervised NLDR technique that utilizes a boosted distance metric (BDM) in place of the EDM. The BDM, which was first introduced,[21] employs the AdaBoost algorithm developed by Freund and Schapire.[22] AdaBoost is an ensemble approach that allows for implicit feature weighting based on class discriminability. The difference between SE and BoSE is that BDM actively places importance on discriminatory features while mitigating the role of weaker features, yielding an embedding which encourages same class objects to be embedded closer to each other and dissimilar class objects to be mapped farther apart. The main purpose of BoSE is to improve the lower-dimensional embedding so that classification and image retrieval will be more accurate in the lower-dimensional spaces. Feature weighting prior to DR makes BoSE a supervised method. Similar to SE, the methods[19,20] are unweighted and unsupervised.

The primary contributions of this work are twofold. First we present a new NLDR scheme boosted spectral embedding (BoSE) that employs AdaBoost with SE to generate lower-dimensional data representations with greater class separability. Second, BoSE is employed in conjunction with a CBIR scheme (CBIR-BoSE)

to perform accurate retrieval of database images with respect to a query instance. An overview of the CBIR-BoSE system is illustrated in Figure 1. For a database of *N* annotated images, feature extraction is performed to yield *N* corresponding high-dimensional feature vectors. A subset of the *N* high-dimensional feature vectors is used as a training set for a boosted classifier to compute the weights for each feature in this two-class CBIR problem. A low-dimensional embedding of the entire dataset (M^{BoSE}) is then created via BoSE. The Euclidean distance between the query image and the database images is then computed in M^{BoSE} and the most similar (lowest distance) images are first retrieved. Images retrieved from the same class as the query instance are considered as "relevant". Evaluation is done by constructing precision-recall (PR) curves, where a large area under precision-recall curves (AUPRC) reflects that CBIR-BoSE is retrieving the most relevant images first.

In this work we evaluated our CBIR-BoSE system on three different two-class problems, illustrated in Figure 2. The three datasets comprised (1) 58 hematoxylin and eosin (H and E) stained prostate cancer tissue biopsy samples classified as benign [Figure 2a] or malignant [Figure 2d] and these images were represented by Gabor, Haralick, and first-order statistical features;

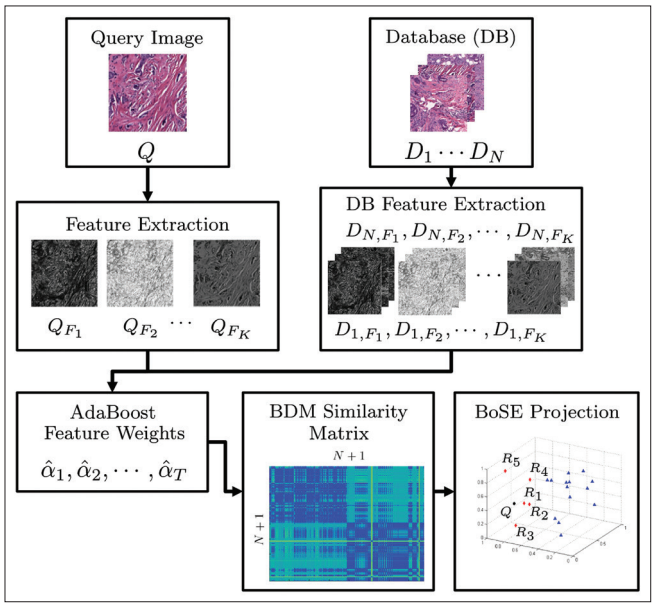

Figure 1: A flowchart illustrating the different components of the content-based image retrieval/boosted spectral embedding (CBIR-BoSE) system. Initially quantitative feature extraction is performed on a query image Q to yield a set of K image descriptors F1,....Fk. The database contains N annotated images (with corresponding class labels) with their corresponding feature-based representations. For the particular problem of interest, the image features are assigned weights ($\hat{\alpha}_1,\ldots,\hat{\alpha}_T$) corresponding to their class separability. A weighted similarity matrix is then created via the boosted distance metric, which is then used with BoSE to project the data into a lower-dimensional space. In the reduced space, the distance between the query Q and the database images is calculated and the database images most similar to the query are retrieved (R1, ..., R5)

(2) 55 H and E stained estrogen receptor (ER) + breast cancer histology specimens classified as low [Figure 2b] or high [Figure 2e] grade and these images were represented by Haralick features; and (3) 41 H and E stained HER2+ breast cancer tissue specimens classified as having low [Figure 2c] or high [Figure 2f] levels of lymphocytic infiltration (LI) and these images were represented by architectural features using the delaunay triangulation, minimum spanning tree, and the voronoi diagram. The choice of these datasets was dictated by the fact that manual inspection of both prostate and breast cancer histology suffers from high inter- and intra-pathologist variability.[23-25] Typically the pathologist first determines if the histology sample is benign or malignant. If it is found to be malignant, the cancer is assigned a grade based on the morphologic and architectural attributes; cancer grade being highly correlated to patient outcome[23,26] In the progression of solid tumors, local and systemic inflammation tends to play an important role.[27] Tumor infiltrating lymphocytes represent a local immune response and the degree of LI in a tumor is considered as being prognostic of patient outcome in several different disease states.[28-30] The development of CBIR tools with applications in digital pathology[31] could assist pathologists by providing a quantitative, reproducible and accurate image-based risk score, indicative of disease aggressiveness and patient outcome.[23] Additionally, a CBIR system for digitized histopathology could serve as a teaching, training, and instructional tool for pathology residents and fellows.

The rest of this paper is organized as follows. The BDM is presented in Section 2. The methodological description of the BoSE scheme is presented in Section 3. The experimental design and evaluation of BoSE are presented in Section 4. Results and discussion are presented in Section 5. Lastly, concluding remarks are presented in Section 6 [Table 1].

BOOSTED DISTANCE METRIC

Brief Overview of Boosted Distance Metric

We define a set of objects as $X = \{x_1, x_2,\ldots, x_N\}$ where N is the number of objects and each image represents one object. Each image x_i, $i \in \{1,\ldots, N\}$ belongs to one of two classes + 1 or − 1. The ground truth label of x_i is denoted $L(x_i) \in \{+1, -1\}$ where $L(x_i) = -1$ indicates membership in class − 1 and $L(x_i) = 1$ indicates membership in class + 1. Let $\Phi_d(x_i)$ for $d \in \{1, 2,\ldots, D\}$ represent the value of feature d from x_i. The BDM construction is comprised of three main steps:

Step 1: Constructing weak classifiers: Weak classifier $h_d(x_i) \in \{-1, 1\}$ predicts the class label of x_i based on feature operator Φ_d. In this work, a weak classifier is one that outputs a class label for the object under consideration. The weak learner may be one that outputs a probabalistic

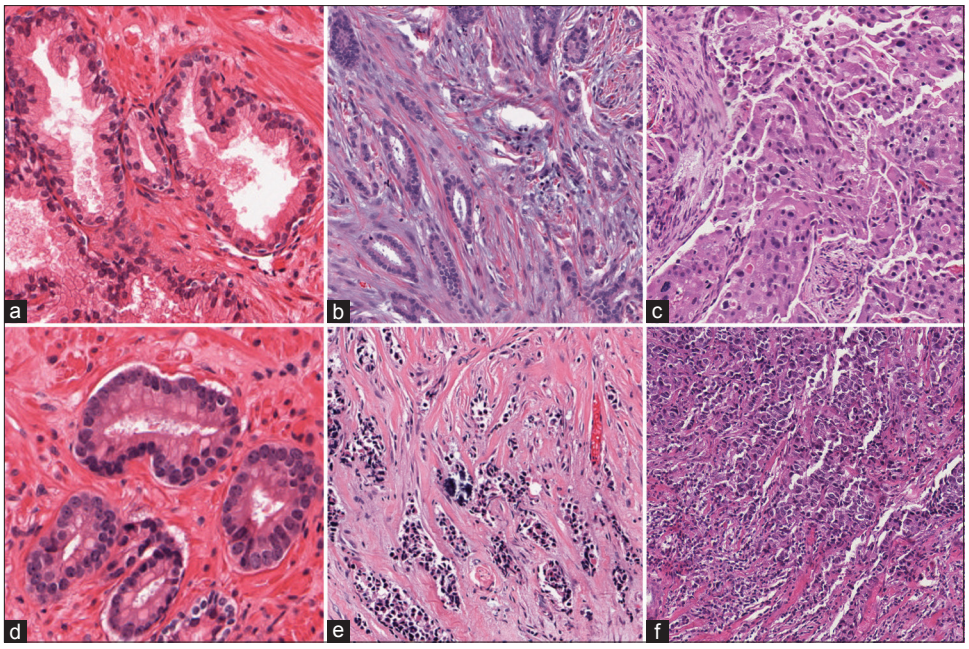

Figure 2: Example images of (a) benign and (d) malignant prostate tissue, (b) low and (e) high grade estrogen receptor (ER+) breast cancer tissue, and HER2+ breast cancer tissue with (c) low and (f) high levels of lymphocytic infiltration. The histology images were obtained by digitizing biopsy samples previously stained with hematoxylin and eosin (H&E). In (a) the nucleoli are less prominent and the glands are more open, whereas in (d) the nucleoli are more apparent and the glands are shriveled due to increased cell proliferation. There is a greater amount of nuclear proliferation in (e) high grade ER+ breast cancer when compared to (b). A similar phenomenon can be observed when looking at HER2+ breast cancer tissue with low versus high levels of lymphocytic infiltration. In (f) there are more lymphocytes that have infiltrated the cancerous tissue compared to (c)

likelihood that an object (in this case, an image) belongs to a specific class based solely on a single attribute. These probabilities can be thresholded to obtain the class label. Multiple different weak learners derived from various image features can be constructed and evaluated in terms of classifier accuracy (assuming that a training set with class labels is available). Weak classifiers were constructed by using only a subset (training set) of the entire dataset

Step 2: Implicit Feature Weighting: The T most accurate weak classifiers, h_t, $t \in \{1, 2,\ldots, T\}$ are identified and weights $\hat{}$ associated with each h_t are learned via the AdaBoost[22] algorithm, thereby enabling implicit feature weighting

Step 3: BDM Construction: The BDM is then defined using the features $\Phi_t(x_i)$ and associated weights $\hat{\alpha}_t$ obtained in Step 2.

Construction of Weak Classifiers

Each individual feature (weak classifier) is used to classify an image and its classification accuracy is leveraged in determining its class separability. The construction of the weak classifiers employed in this work is outlined below:

Step 1: Calculate $\Phi_d(x_i)$ for all $d \in \{1, 2,\ldots, D\}$, $i \in \{1, 2,\ldots, N\}$, in order to obtain corresponding feature values for each of the images.

Step 2: Create training set $X^{tr} \subset X$ containing N objects by randomly sampling half of the entire dataset X

Step 3: Let X + indicate all objects in X^{tr} belonging to class + 1. Similarly, X − is the set of all samples in X^{tr} that belong to class − 1. We can obtain an appropriate probability distribution function (PDF), the integral of the density function, which predicts the likelihood of observing a feature value given a class label as:

$$p(\Phi_d(X_a)\,|\,\omega_b) = \Phi_d(X_a)^{\tau-1}\,\frac{\exp\left(\frac{-\Phi_d(X_a)}{\eta}\right)}{\eta^{\tau}\Gamma(\tau)} \qquad (1)$$

for $a \in \{+,-\}$, $\omega_b \in \{+1,-1\}$ Γ is the gamma function, and $\tau, \eta > 0$ are scale and shape parameters. Equation 1 is a gamma function estimation of the PDF,[32] and is preferred to a Gaussian distribution because the feature histograms are asymmetric about the mean and the gamma function models the distribution more accurately.

Step 4: Obtain the *a posteriori* probability $P(+1\,|\,\Phi_d(x_i))$ which computes the likelihood that an object with feature value $\Phi_d(x_i)$ belongs to the positive class + 1 by solving,

$$P(+1\,|\,\Phi_d(x_i)) = \frac{P(+1)p(\Phi_d(x_i)\,|\,+1)}{P(+1)p(\Phi_d(x_i)\,|\,+1)+ P(-1)p(\Phi_d(x_i)\,|\,-1)} \qquad (2)$$

Step 5: Once the *a posteriori* probabilities have been computed for each image based on a single

feature, the weak classifiers are defined based on the individual features. The weak classifiers may now be defined as follows:

$$h_d(x_i) = \begin{cases} 1 & \text{if } P(+1|\Phi_d(x_i)) > P(-1|\Phi_d(x_i)) \\ -1 & \text{otherwise} \end{cases}$$

If the probability, which based on a single feature, of the image x_i belonging to class $+1$ is greater than its probability of belonging to class -1, it will be given a class label of 1. Otherwise, it will be given a classification label of -1.

Implicit Feature Weighting

We use the AdaBoost[22] algorithm to perform implicit weighting of the weak classifiers (in turn reflecting the importance of the individual image attributes) in order to distinguish between the positive and negative classes. Our feature weighting algorithm is illustrated in Figure 3. AdaBoost works in an iterative fashion by first identifying the best-performing weak classifiers and then assigning weights based on the discriminative value of that feature.[22] The weights of the training images are initialized by taking the reciprocal of the number of images there are in the training set (Line 0). For each weak classifier (feature), its classification error is computed (Line 2). At each iteration, the weak classifier with the lowest classification error is chosen and its weight is determined (Line 4). The weights of the training images are updated such that the images that were frequently classified properly received lower weights, while the images that were frequently misclassified received higher weights (Line 5). This ensures that subsequent weak classifiers are picked based on their ability to classify these hard to classify instances. The process repeats T for iterations. The output of the algorithm is a set of weak classifiers h_t and their associated normalized weights $\hat{\alpha}_t$, t ∈ {1, 2,..., T} where $1 \leq T \leq D$ and $0 < \hat{\alpha}_t < 1$. $\hat{\Phi}_t$ is the operator for the feature selected at iteration t of AdaBoost. The algorithm stops when $\varepsilon_t > 0.5$.

Constructing the Boosted Distance Metric

The BDM is constructed after the weights and features have been chosen. To find the distance between two points in the high-dimensional space, we calculate,

$$D_{BDM}(x_i, x_j) = \left[\sum_{t=1}^{T} \hat{\alpha}_t \left(\hat{\Phi}_t(x_0) - \hat{\Phi}_{0<<1.}(x_j) \right)^2 \right]^{\frac{1}{2}} \quad (3)$$

This is essentially a weighted Euclidean distance, where the weights influence the contribution of each feature. If $\hat{\alpha}_t \approx 0$, then $\hat{\Phi}_t$ will not affect the value of the similarity measure.

Proposition 2.1 given that $D_{Eu} = \left[\sum_{t=1}^{T} (\Phi_t(x_i) - \Phi_t(x_j))^2 \right]^{\frac{1}{2}}$ is the Euclidean distance metric, is also a distance metric.

Proof since D_{Eu} is a metric, it is (1) positive, (2) symmetric, (3) definite, and (4) the triangle inequality holds. D_{BDM} must also be a metric since $\hat{\alpha}_t \in [0.1]$ is positive and real valued. Therefore properties (1)-(4) are satisfied for D_{BDM}.

A look at the simple case where T = 2, and where a, b ∈ R^2 can provide some insight into D_{BDM}.

If L (a) = L (b) then on average, over the entire training set, D_{Eu} (a, b) > D_{BDM} (a, b). Ideally, if L (a) = L (b), then D (a, b) ≈0. We denote the distance between a and b in the first dimension as Δ_1 and the second dimension as Δ_2. Assume that feature dimension Δ_1 is on average, over the entire training set, more discriminating than Δ^2; more specifically that $||\delta_1$ (a) −

Algorithm: *BoostFeatWeights*
Input: Training samples $\mathbf{X}^{tr}$, ground truth labels $\mathcal{L}(\mathbf{X}^{tr})$, iterations T, weak classifiers h_d for $d \in \{1, 2, \ldots, D\}$
Output: Optimal classifiers h_t and their corresponding weights $\hat{\alpha}_t$
begin
0. Initialize distribution for samples $\Pi_1(i) = \frac{1}{N}$
1. **for** $t = 1$ to T
2. Find $h_t = \arg\min_{h_d}[\epsilon_d]$, where $\epsilon_d = \sum_{i=1}^{N} \Pi_t(i)[\mathcal{L}(\mathbf{x}_i) \neq h_d(\mathbf{x}_i)]$ for $\mathbf{x}_i \in \mathbf{X}^{tr}$;
3. *if* $\epsilon_t \geq 0.5$ *then* stop;
4. $\alpha_t = \frac{1}{2} \ln \frac{1-\epsilon_t}{\epsilon_t}$;
5. Update, $\Pi_{t+1}(i) = \frac{1}{Z_t} \Pi_t(i) \exp(-\alpha_t \mathcal{L}(\mathbf{x}_i) h_t(\mathbf{x}_i))$ for all $\mathbf{x}_i \in \mathbf{X}^{tr}$, where $Z_t = \sum_i \Pi_t(i) \exp(-\alpha_t \mathcal{L}(\mathbf{x}_i) h_t(\mathbf{x}_i))$ is a normalization term;
6. **endfor**
7. Normalize α_t to obtain $\hat{\alpha}_t$ such that $0 < \hat{\alpha}_t \leq 1$, $\hat{\alpha}_t = \frac{\alpha_t}{\max_t[\alpha_t]}$ for $t \in \{1, \ldots, T\}$.
8. **return** $\hat{\alpha}_t$ and h_t;
end

Figure 3: The BoostFeatWeights algorithm for implicitly weighting the top performing image features for a specific task. All samples were initialized with equal weights. The weights for the weak classifiers are computed based on the classification error ε_d. At each iteration, weights $(\Pi_t(i))$ increase for samples that are difficult to classify. This forces the weak classifiers to concentrate on the images that are frequently misclassified. Once all the weights (α_t) for the weak classifiers are found, they are normalized so that they would range from 0 to 1. The T best performing classifiers and their weights are computed

Algorithm: *BoSE*
Input: Training samples $\mathbf{X}^{tr}$, Testing samples $\mathbf{X}^{te}$, $\mathcal{L}(X^{tr})$, $\mathcal{L}(X^{te})$, iterations T
Output: Lower-dimensional embedding $\mathbf{Y}$
begin
1. Build weak classifiers $h_d : d \in \{1, 2, \ldots, D\}$ via a Bayesian Classifier;
2. Select optimal weak classifiers h_t and weights $\hat{\alpha}_t$ for $t \in \{1, 2, \ldots, T\}$ via AdaBoost;
3. Obtain BDM by applying Equation 3;
4. Obtain $\mathbf{W}$ by Equation 9;
5. Find $\mathbf{Y} \in \mathbb{R}^{N \times k}$;
6. **return** $\mathbf{Y}$

end

Figure 4: The BoSE algorithm. The weak classifiers are built using the training samples ($\mathbf{X}^{tr}$) and the weights are calculated via AdaBoost. The boosted distance metric is then employed with the weights to calculate the distances between all the objects in X. The distances are used in conjunction with the Gaussian kernel to obtain the weight matrix W. The lower-dimensional embedding Y is then obtained by solving the eigenvalue decomposition in Equation 8

δ_1 (b) $||<||$ δ_2 (a) $-\delta_2$ (b)$||$ where δ_1 and δ_2 represent the positions of the objects in feature spaces Δ_1 and Δ_2 respectively. Thus, $\hat{\alpha}_1 > \hat{\alpha}_2$ via the learned feature weights. Recall that $D_{BDM}(a,b) = \sqrt{\hat{\alpha}_1(\Delta_1)^2 + \hat{\alpha}_2(\Delta_2)^2}$ and $D_{Eu}(a,b) = \sqrt{(\Delta_1)^2 + (\Delta_2)^2}$. It can be seen that on average, over the entire training set, the following holds:

$$\sqrt{(\Delta_1)^2 + (\Delta_2)^2} > \sqrt{\hat{\alpha}_1(\Delta_1)^2 + \hat{\alpha}_2(\Delta_2)^2} \quad (4)$$

$$(\Delta_1)^2 + (\Delta_2)^2 > \hat{\alpha}_1(\Delta_1)^2 + \hat{\alpha}_2(\Delta_2)^2 \quad (5)$$

$$(\Delta_1)^2 - \hat{\alpha}_1(\Delta_1)^2 > \hat{\alpha}_2(\Delta_2)^2 - (\Delta_2)^2 \quad (6)$$

$$(\Delta_1)^2(1 - \hat{\alpha}_1) > (\Delta_2)^2(\hat{\alpha}_2 - 1) \quad (7)$$

Recall that $\hat{\alpha}_1, \hat{\alpha}_2 \geq 0$ and $\hat{\alpha}_1, \hat{\alpha}_2 \in [0,1]$. Therefore, the left hand side of the inequality would yield a positive number and the right hand side would yield a negative number. Therefore on average, over the entire training set, $D_{Eu}(a,b) < D_{BDM}(a,b)$ if $L(a) = L(b)$. Note that it is similarly possible to show that under the same assumptions made for proposition 2.1 if $L(a) \neq L(b)$, then on average, over the entire training set, $D_{Eu}(a,b) < D_{BDM}(a,b)$.

BOOSTED SPECTRAL EMBEDDING FOR CONTENT-BASED IMAGE RETRIEVAL

Boosted Spectral Embedding

The goal of SE is to project the feature vectors from a D dimensional space to a k dimensional space, where k << D. The low-dimensional representation of X is denoted $Y = \{y_1, y_2, \ldots, y_N\}$. The first step in SE is to create a weight matrix W where each element (i, j) in W is denoted by w_{ij} and represents the distance between x_i and x_j defined by some metric D.

The low-dimensional representation of X is then found by solving the eigenvalue decomposition problem:

$$(L - W)Y = \lambda LY, \quad (8)$$

where L is the diagonal matrix, $L_{jj} = \sum_i w_{ij}$ (15)

The typical formulation of W involves the use of the EDM, where $w_{ij} = \exp(-D_{Eu}(x_i, x_j)/\sigma)$, and σ is the standard deviation of X. However, in BoSE, we replace the EDM with the BDM to obtain,

$$w_{ij} = \exp\left(-\frac{D_{BDM}(x_i, x_j)}{\sigma}\right) \quad (9)$$

Since SE seeks to preserve object adjacencies as defined by W by improving the description of adjacency via the BDM, we should improve the resulting low-dimensional embedding (achieve greater class separability in the reduced embedding space). Since D_{BDM} is a metric, W is positive, semi-definite, and symmetric.

Content-Based Image Retrieval - Boosted Spectral Embedding

The high-dimensional feature data extracted from each of the datasets is reduced to a fewer number of dimensions via BoSE, the intent being to perform retrieval in the BoSE reduced space. Briefly, the retrieval is performed as follows. The query sample and all existing annotated database samples are aggregated and the BoSE representation for all images (following feature extraction and weighting) is determined. Using the EDM, the distance between the query image and all of the database images is calculated in the BoSE space. The resulting distance vector is sorted in ascending order and the most similar database images in terms of distance are outputted. The CBIR-BoSE algorithm is illustrated in Figure 5.

Experimental Design and Evaluation

Dataset Description

We considered three datasets [Table 2]. Slides from all three datasets were stained with H and E and scanned into a computer via a whole-slide digital scanner at the University of Pennsylvania (prostate cancer) and the Cancer Institute of New Jersey (breast cancer). The prostate and breast cancer images were taken at magnifications of ×40

and ×20, respectively, and were saved in the SVS format. Pathologists were instructed to manually place a contour around homogeneous regions of tissue corresponding to either "cancer" or "noncancer" regions. Annotation was performed on the scanned SVS biopsy image files using the Image Scope software platform (Aperio ePathology, Leica Biosystems). No confounding tissue types (e.g., atrophy, prostatic intraepithelial neoplasia) were included. The entire set of tissue biopsy images were then divided into 30-by-30 square pixel regions; within these 900 pixels, if over 50% of the pixels (450) contained the expert's annotation, those regions were included in the dataset. All of the images were converted from the RGB color space to the hue, saturation, value (HSV) space to mitigate the effect of varying stain intensities. By converting images to the HSV space, we ensure that any potential stain intensity variation across images is confined to a single channel (the "value" channel). All three channels are still evaluated. However, if the variation in the intensity is detrimental, the features in the "value" channel will not be selected. The objective of experiment 1 (D_1) was to distinguish between malignant and benign prostate tissue patches [Table 2, Figure 6] from biopsy samples obtained from 58 patients. In experiment 2 (D_2), we aimed to distinguish between high and low grade breast cancer tissue patches from biopsy samples obtained from 55 patients. Lastly, the objective of Experiment 3 (D_3) was to distinguish between high and low levels of LI in breast cancer tissue patches from 41 biopsy samples obtained from 12 patients. The final diagnosis and grade for each of the datasets was obtained as a consensus of two expert pathologists.

Content-Based Image Retrieval Comparisons

For each of the experiments detailed below, we compared the performance of three CBIR paradigms.

Algorithm: *CBIR-BoSE*
Input: Query image $\mathbf{Q}$, database images $\mathbf{X}^{db} \in \mathbb{R}^{N \times D}$
Output: Top $\mathcal{N}$ Retrieved Images
begin
1. Calculate $\mathbf{x}^{query} = \Phi_d(\mathbf{Q})$ for all $d \in \{1, 2, \ldots, D\}$, where $\mathbf{x}^{query} \in \mathbb{R}^{1 \times D}$;
2. Concatenate $\mathbf{x}^{query}$ with $\mathbf{X}^{db}$ to form $\mathbf{X}^{all} \in \mathbb{R}^{(N+1) \times D}$;
3. Input $\mathbf{X}^{all}$ into BoSE to yield $\mathbf{Y}^{all} \in \mathbb{R}^{(N+1) \times k}$ where $k << D$;
4. Extract reduced query vector from $\mathbf{Y}^{all}$ to yield $\mathbf{y}^{query} \in \mathbb{R}^{1 \times k}$ and $\mathbf{Y}^{db} \in \mathbb{R}^{N \times k}$;
5. Calculate $\mathbf{p} = \mathbb{D}_{\text{Eu}}(\mathbf{y}^{query}, \mathbf{Y}_i^{db}), i \in \{1, 2, \ldots, N\}, \mathbf{p} \in \mathbb{R}^{1 \times N}$;
6. Rearrange $\mathbf{p}$ in ascending order from the smallest to the largest value.
7. Extract the $\mathcal{N}$ smallest values and find the corresponding images.
8. **return** $\mathcal{N}$ most similar images.

end

Figure 5: The content-based image retrieval/boosted spectral embedding algorithm

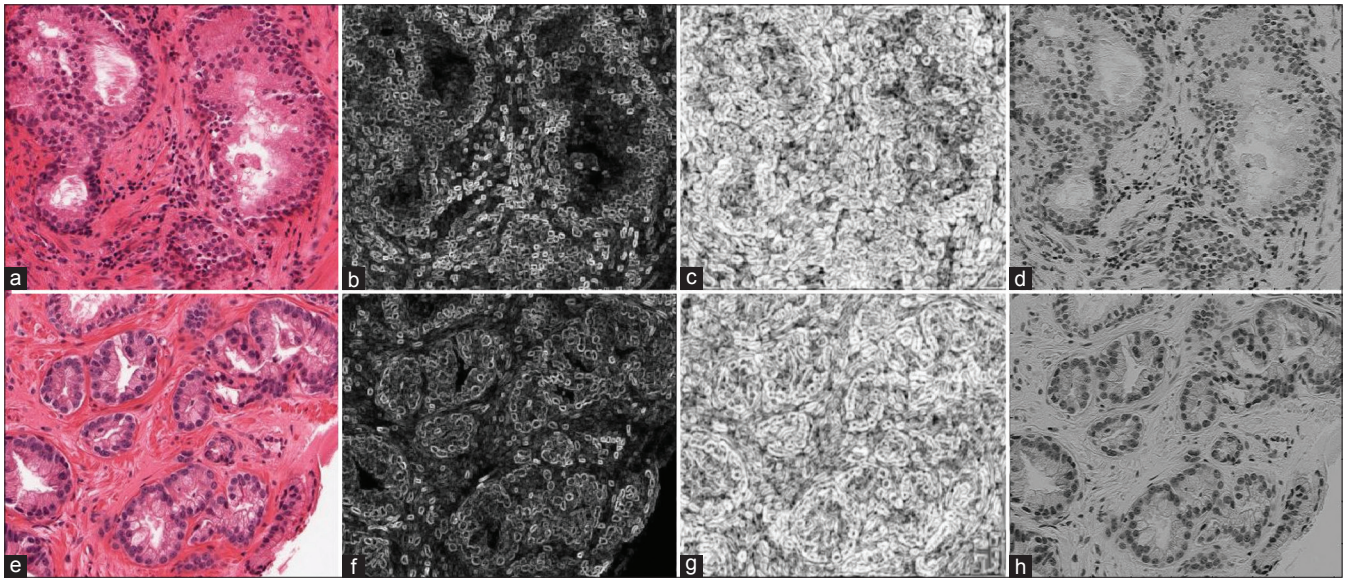

Figure 6: Examples of (a) benign and (e) Gleason grade 3 prostate cancer images and their corresponding feature images obtained at the level of sub-regions of the whole image: (b) (f) First-order statistics (range using a 5 × 5 window, Hue color channel), (c) (g) Haralick (Correlation using a 5 × 5 window, Hue color channel), and (d) (h) Gabor features (5 × 5 window, Hue color channel)

(1) CBIR-BoSE: The proposed system, whereby the BDM is used to perform SE and obtain a low-dimensional representation of objects for retrieval [Figure 5]. (2) CBIR-SE: Traditional SE, whereby the EDM is used to obtain the low-dimensional representation. (3) CBIR-BDM: CBIR performed using BDM to identify object similarities in the original high-dimensional space. These three paradigms are evaluated in terms of CBIR measures of performance. CBIR-BoSE and CBIR-SE are also evaluated in terms of classification accuracy. These are detailed below in Section 4.6 "Evaluation".

Experiment 1: Distinguishing Malignant from Benign Prostate Histopathology

In,[32] Doyle, *et al.* found that a weighted combination of Gabor filter features, Haralick co-occurrence features, and first-order statistics could accurately distinguish between benign and malignant patterns of H and E-stained prostate tissue. In malignant tissue cell proliferation is increased, leading to enlarged nuclei and visible nucleoli, which cause malignant tissue to absorb an increased amount of hematoxylin. These changes lead to clear texture differences [Figure 6], which are captured by the following features:

- Gabor filter features:[33] Gabor filter banks are created by modulating a Gaussian function by a sinusoid, parameterized by an orientation and a frequency parameter. The filter provides a large response when convolved with directional image intensity patterns that correspond to the filter's parameters. The increased size and number of nuclei, in addition to the appearance of clear nucleoli, cause significant changes in the response at small frequency parameters
- Haralick co-occurrence features: [34] Co-occurrence image features are based on the adjacency of pixel values in an image. An adjacency matrix is created where the value of the ith row and the jth column equals the number of times pixel values i and j appear within a fixed distance of one another. By calculating these feature values for a local neighborhood, we can discriminate between the different texture patterns produced by benign and malignant nuclei
- First-order statistical features: Simple statistics calculated on image values are used to quantify intensity variations. These texture features include the mean, median, and standard deviation of local neighborhoods as well as gradient features and directional derivatives, which indicate transitions between high intensity values (the stroma/lumen) and the low intensity values (nuclei and nucleoli). These changes should be pronounced in malignant tissue, where nuclei are stained more heavily.

Feature operators extract these three families of quantitative features from each channel in an image, yielding a total dataset of over 900 features. In,[32] 14 highly discriminating pixel-wise features were learned via AdaBoost[22] out of a feature set that comprised over 900 features. AdaBoost assigned a weight to all of the features and these weights were thresholded in that features with a $\alpha > 0.05$were retained while the noninformative features were discarded. In the current study these 14 features were extracted for each image, generating 14 corresponding feature images, three of which are illustrated in Figure 6. The pixel values for each feature image were averaged, generating a 14 element feature vector to characterize each prostate image [Table 3].

Experiment 2: Distinguishing High from Low Grade Breast Histopathology

Two of the defining histological features of breast cancer are the disorganization of the tissue and the structure of the cells. The severity of the cancer is given a Bloom Richardson (BR) grade level.[26] Breast cancer tissue samples with greater disorganization and increasingly irregular structure are given higher grades. High grade samples exhibit more nuclear proliferation than low grade samples. As with the prostate cancer samples, the breast cancer biopsy samples were stained with H and E. Haralick features were extracted and used to describe

Table 1: List of mathematical symbols and notations used throughout the paper

Symbol	Description
$X = \{X_1, X_2, ..., X_N\}$	Quantitative representation of images in $R^{N\times D}$
$Y = \{y_1, y_2, ..., y_N\}$	Low-dimensional projection of X
W	Weight matrix
Φ_d	Feature operator that extracts quantitative feature d from image
$L(x_i) \in \{+1, -1\}$	Ground truth label for object x_i
h_d	Weak classifier built using a Bayesian framework
$\alpha_t, t \in \{1,....,T\}$	Weights associated with the T most optimal features
$\hat{\alpha}_t, t \in \{1,....,T\}$	Normalized weights associated with the T most optimal features
D_{BDM}	Boosted distance metric
M^{BOSE}	Low-dimensional representation produced by BoSE
M^{SE}	Low-dimensional representation produced by SE

BoSE: Boosted spectral embedding, SE: Spectral embedding

Table 2: List of the breast cancer and prostate cancer datasets used in this study

Data	Classes (+1/−1)	Class distribution (+1/−1)	Number of samples
Prostate (D_1)	Cancer/benign	29/29	58
Breast (D_2)	High grade/low grade	36/19	55
Breast (D_3)	High LI/low LI	20/21	41

LI: Lymphocytic infiltration

Table 3: Texture features extracted from the prostate tissue sample images

Texture feature	Parameters	Total features
First-order statistics (SD, range)	Window size: $w = 5$	2
Haralick features (information measure, correlation, energy, contrast variance, entropy)	Window size: $w = 5$ Distance: $\delta = 1$	5
Gabor features	Window size: $w \in \{5, 9\}$ Orientation: $\theta \in \{0, \frac{\pi}{6}, \ldots, \frac{5\pi}{6}\}$	7

SD: Standard deviation

the degree of nuclear proliferation by quantifying the variations in the intensity values in the images. The objective of this experiment was to retrieve images corresponding to the grade of the query image. To define a two-class problem, all images are first separated into either low (BR 4, 5) and high (BR 7, 8) grade classes [Table 2]. From each image, 12 Haralick feature images were generated (contrast energy, contrast inverse moment, contrast average, contrast variance, contrast entropy, intensity average, intensity variance, intensity entropy, entropy, energy, correlation, and one information measure of correlation) and the following statistics were computed from the pixel values from each feature image: mean, standard deviation, and entropy. This was done for all three color channels in the HSV space.

Experiment 3: Distinguishing High Lymphocytic Infiltration from Low Lymphocytic Infiltration Breast Histopathology

The class problem is defined as follows: Images were separated into either low LI or high LI classes [Table 2]. To quantify the arrangement of lymphocytic nuclei in the histology images, architectural features were computed for each image. The centroids of the lymphocytic nuclei are used to construct the delaunay triangulation G_D [Figure 7b and f], the minimum spanning tree G_M [Figure 7c and g], and the Voronoi Diagram G_V [Figure 7d and h]. Automated nuclear detection was performed to identify the nuclear centers as centroids of the different graphs. However, the cancer and lymphocytic nuclei are similar in appearance. In general, lymphocytic nuclei differ in appearance from cancer cell nuclei by their smaller size, more circular shape, and a darker homogeneous staining.[35] We took these differences into account and performed automated nuclear detection in the following manner.

Step 1: On each image, M candidate nuclear centers $M = \{m_1, m_2, \ldots, m_M\}$ were found by convolving the image x_i with a Gaussian (smoothing) kernel at multiple scales. This was done to account for the variation in lymphocyte size. The darkest pixels were found on the smoothed image based on local differences in luminance and these were the candidate lymphocytic nuclear centers.

Step 2: Using Hojjatoleslami and Kittler's region-growing scheme,[36] each of the M candidate lymphocytic nuclear centers was grown into a corresponding region . The optimal regions were identified when the boundary strength, which is defined as the difference in the mean intensity of the pixels in the internal boundary and the current boundary of the region, was at a maximum. See[35] for a more detailed description.

Step 3: Each of $r \in R$ contained two random variables: $A_r \in \{\omega_c, \omega_l\}$ which is the classification of the candidate nuclear centers as either a cancer (ω_c) or lymphocytic (ω_l) nucleus and $B_r \equiv [C_r, \phi_r]^T \in R(^{+2})$ where C_r is the square root of the nuclear area and ϕ_r is the standard deviation of the luminance in the nuclear region. The labels, A_r, given the feature vectors B_r are estimated via a maximum *a posteriori* (MAP) estimation by finding the A_r that maximizes the posterior probability.

$$p(A_r \mid B_r) = \frac{p(B_r \mid A_r)p(A_r)}{p(B_r)} \quad (10)$$

where $p(B_r|A_r)$ is the likelihood term and $p(A_r)$ and $p(B_r)$ are prior probabilities. $p(B_r)$ is ignored because maximization was done with respect to $p(A_r)$.

Step 4: $p(B_r|A_r)$ is computed from PDFs, where A_r is provided by manual delineation of lymphocytes in a training set.

Step 5: The prior probabilities $p(A_r)$ is defined by a Markov random field) andcomputed. The iterated conditional modes algorithm,[37] a deterministic relaxation procedure, was used to compute the MAP estimation and classify each $r \in R$. The regions classified as cancer nuclei were discarded and the centroids of the lymphocytic nuclei were calculated, yielding $O=\{o_1, o_2, \ldots, o_L\}$ where $O \subseteq M$. Details of the automated nuclear detection can be found in.[35]

Using the O centroids, we constructed a graph G = (V, E, J), where V represent the vertices of the graph which correspond to the number of centroids, E are the set of edges, and J are the weights of the edges, proportional to edge length. The set of vertices, edges, and weights make up a unique graph on the image. From each graph, we extracted a set of features listed in Table 4. A detailed description of the graph construction and feature extraction can be found in.[38]

EVALUATION

Content-Based Image Retrieval - Boosted Spectral Embedding

The performance of a CBIR system is determined by how many retrieved images for a given query image

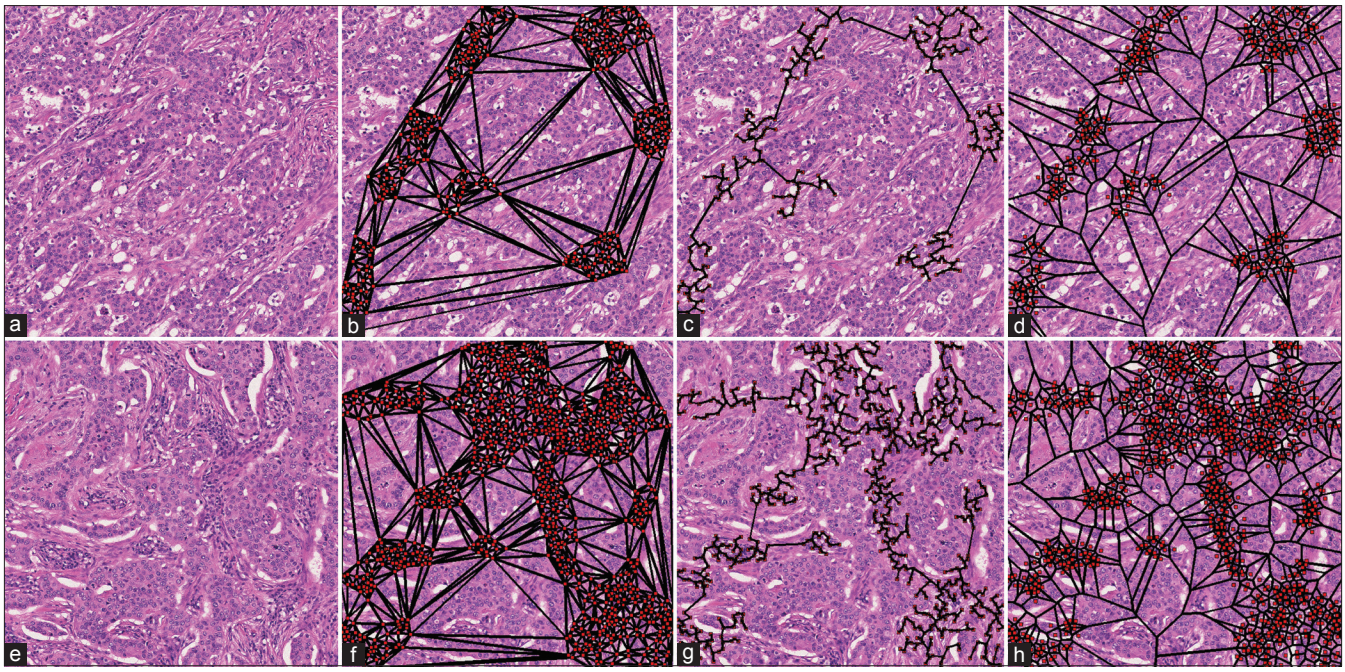

Figure 7: Example breast histopathology images that contain (a) low and (e) high levels of lymphocytic infiltration with their corresponding feature images: (b) (f) Delaunay Triangulation, (c) (g) Minimum Spanning Tree, and (d) (h) Voronoi Diagram. Quantitative graph features were calculated using the graphs constructed on the image

Table 4: List of the features extracted to quantify the degree of LI. A detailed description of the feature extraction and graph construction can be found in[38]

Graph	Features
Voronoi diagram (13 features)	Total area of all polygons
	Polygon area (mean, SD, minimum/maximum ratio, entropy)
	Polygon perimeter (mean, SD, minimum/maximum ratio, entropy)
	Polygon chord length (mean, SD, minimum/maximum ratio, entropy)
Delaunay triangulation (8 features)	Triangle area (mean, SD, minimum/maximum ratio, entropy)
	Triangle side length (mean, SD, minimum/maximum ratio, entropy)
Minimum spanning tree (4 features)	Branch length (mean, SD, minimum/maximum ratio, entropy)
Nuclear features (25 features)	Density of nuclei
	Distance to {3, 5, 7} nearest nuclei (mean, SD, disorder)
	Number of nuclei in a {10, 20, ..., 50} pixel radius (mean, SD, disorder)

LI: Lymphocytic infiltration, SD: Standard deviation

are relevant to the query, defined as images which belong to the same class as the query image, and also the order in which they appear. Precision is defined as $\rho(\beta)=\frac{\xi(\beta)}{\beta}$, where $\xi(\beta)$ denotes the number of relevant objects in the β closest objects. Recall is defined as $r(\beta)=\frac{\xi(\beta)}{\xi(N-1)}$. Precision-recall curves were generated by plotting $p(\beta)$ versus $r(\beta)$ for $\beta\in\{1,2,...,N-1\}$. Area under the AUPRC was measured and used to evaluate the CBIR system. The AUPRC $\in [0,1]$values where an AUPRC 1 indicates that the CBIR system only retrieved relevant images and an AUPRC 0 indicates that the CBIR system only retrieved irrelevant images. Therefore, the higher the AUPRC, the better the CBIR system. We denote θ_{BoSE}^{AU}, θ_{SE}^{AU} and θ_{BDM}^{AU} as the AUPRC values for CBIR-BoSE, CBIR-SE, and CBIR-BDM, respectively. CBIR-BDM retrieves images from the database using the BDM without DR.

Classifier Evaluation of Boosted Spectral Embedding and Spectral Embedding

A second performance measure for evaluating BoSE is classifier accuracy. Of the classifiers available (Support Vector Machines, Neural Nets, etc.), the Random Forest (RF) classifier was chosen due to its ability to accurately and efficiently run on large databases with minimal training time and lower overall computational time. The RF classifier (obtained by bagging decision trees)[39] is trained on both $M^{\text{BoSE}}$and M^{SE} [Figure 10].

The accuracy of the RF classifier should reflect the class discriminability of M^{BoSE}. A RF classifier is an ensemble of decision trees (i.e., weak learners) combined via bootstrap aggregation. Averaging decisions across weak learners creates a strong learner that reduces overall bias and variance.[39] We define θ_{BoSE}^{Acc} and θ_{SE}^{Acc} as the classification accuracy when performing classification in the lower-dimensional spaces created by BoSE and SE, respectively. The classification accuracy is defined as $\frac{TP+TN}{TP+TN+FP+FN}$ where *TP* are the true positives, *TN* are the true negatives, FP are the false positives, and FN are the false negatives.

Let $S_{+1} \subset X$ and where for any S_{+1}, $L(a) = +1$ and for any $b \in S_{-1}$, $L(b) = -1$. S_{+1} and S_{-1} are subsets of the total number of the specific class objects we have in X. S_{+1} and are S_{-1} randomly sampled with replacement from X, ensuring that each of S_{+1} and S_{-1} only comprise of instances from either of $+1$and -1. Each random sampling of S_{+1} and S_{-1} is used to train a decision tree classifier Ω_v, where $v \in \{1,2,...,V\}$ and so that $\Omega_v(x) \in \{+1,-1\}$.

Randomized, 3-fold cross-validation was used to determine training and testing inputs for the RF classifier. First, the entire dataset X was randomly divided into three equally-sized subsets X^1, X^2, $X^3 \subset X$. Two of the subsets were used for training the RF classifier, which was then evaluated on the remaining subset. The subsets were subsequently rotated until each subset was used for evaluation exactly once. The entire cross-validation scheme was repeated over 50 iterations, over which mean and standard deviation classification accuracy were reported.

Evaluating Intrinsic Dimensionality for Content-Based Image Retrieval - Boosted Spectral Embedding

When performing retrieval and classification in the lower-dimensional space, identifying the optimal number of dimensions within which to embed the data is a nontrivial task. Each dataset possesses an intrinsic dimensionality in which the classification accuracy and the retrieval performance will be optimal. In order to evaluate the effect of the total number of embedding dimensions to be considered, for the purpose of maximizing classification accuracy and the AUPRC, each dataset was reduced to lower-dimensional embeddings. The corresponding number of dimensions associated with these reduced-dimensional embeddings was varied and BoSE was evaluated in these different spaces [Table 5]. We define $\theta_{BoSE,k}^{Acc}$ and $\theta_{BoSE,k}^{AU}$ as the accuracy and AUPRC using BoSE in k dimensions, where $k \in \{1,2,...,K\}$ and similarly $\theta_{SE,k}^{Acc}$ and $\theta_{SE,k}^{AU}$ for SE The maximum, minimum, and average AUPRC and classification accuracy is reported and calculated in the following manner: $\theta_{\nu}^{\mu,max} = \max_{k}[\theta_{\nu,k}^{\mu}]$, $\theta_{\nu}^{\mu,min} = \min_{k}[\theta_{\nu,k}^{\mu}]$ $\psi_{\nu}^{\mu} = \frac{1}{K}\sum_{k=1}^{K}\theta_{\nu,k}^{\mu}$ where $\mu \in \{Acc, AU\}$ and $\nu \in \{BoSE, SE\}$.

RESULTS AND DISCUSSION

Experiment 1: Distinguishing Malignant from Benign Prostate Histopathology

Quantitative evaluation

Figure 8 and Table 8 reveal that over a range of dimensions, CBIR-BoSE consistently outperforms CBIR-SE in terms of (a) AUPRC, and (b) accuracy. For D_1, and $\theta_{BoSE}^{AU,max}$ were greater than $\theta_{SE}^{AU,max}$ and $\theta_{SE}^{AU,min}$ [Table 6]. The average AUPRC for CBIR-BoSE (ψ_{BoSE}^{AU}) across the all the dimensionalities evaluated was greater than the average AUPRC for CBIR-SE (ψ_{SE}^{AU}) [Table 6]. $\theta_{BoSE}^{Acc,max}$ and $\theta_{BoSE}^{Acc,min}$ were greater than $\theta_{SE}^{Acc,max}$ and $\theta_{SE}^{Acc,min}$ [Table 8]. ψ_{BoSE}^{Acc} was greater than ψ_{SE}^{Acc} and unlike the AUPRC values, the accuracy values remain relatively invariant to the number of dimensions that D_1 is embedded into via BoSE and SE.

Table 5: The original dimensionality of the datasets and their reduced dimensionality employed for evaluating CBIR-BoSE and CBIR-SE. Both CBIR systems were evaluated after projecting the original high-dimensional data into spaces of progressively different reduced dimensions

Dataset	Original dimensionality	Reduced dimensionality
Prostate cancer	14	1, 2, 3, 4, 5, 6, 7
Breast cancer grading	108	1, 2, 3, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50
Lymphocytic infiltration	50	1, 2, 3, 5, 10, 15, 20, 25

CBIR-BoSE: Content-based image retrieval-boosted spectral embedding, CBIR-SE: Content-based image retrieval-spectral embedding, CBIR: Content-based image retrieval

Table 6: Quantitative results showing the maximum, minimum, and mean AUPRC values for Experiment 1 (D_1), Experiment 2 (D_2), and Experiment 3 (D_3). ψ_{BoSE}^{Acc} is greater than ψ_{SE}^{Au} for D_1, D_2, and D_3 and is statistically significant using a $P<0.0.5$

Dataset	$\theta_{BoSE}^{AU,max}$	$\theta_{SE}^{AU,max}$	$\theta_{BoSE}^{AU,min}$	$\theta_{SE}^{AU,min}$	ψ_{BoSE}^{AU}	ψ_{SE}^{AU}
D_1	0.87	0.68	0.70	0.60	0.79	0.63
D_2	0.90	0.90	0.73	0.57	0.79	0.68
D_3	0.75	0.78	0.45	0.36	0.54	0.44

AUPRC: Area under precision-recall curves

Qualitative evaluation

For each of the top five images retrieved, CBIR-BoSE yielded more relevant images compared to CBIR-SE [Figure 9] reflecting that objects from the same class are mapped closer to each other in M^{BoSE}. Figure 10a and d display M^{SE} and M^{BoSE}, respectively, showing a much greater separation between the malignant and benign classes in M^{BoSE} compared to M^{SE}.

Experiment 2: Distinguishing High from Low Grade Breast Histopathology

Quantitative evaluation

For D_2, θ^{AU}_{BoSE} and θ^{AU}_{SE} decreased as the dimensionality of the data increased [Figure 11]. While $\theta^{AU,max}_{BoSE}$ and $\theta^{AU,max}_{SE}$ occurred when D_2 was reduced to two dimensions and were similar, θ^{AU}_{SE} decreased more drastically compared to θ^{AU}_{BoSE}. This resulted in $\theta^{AU,min}_{BoSE}$ being greater than $\theta^{AU,min}_{SE}$ [Table 6]. Another consequence of the difference in the rate of decrease of θ^{AU} between CBIR-BoSE and CBIR-SE was that ψ^{AU}_{BoSE} was greater compared to ψ^{AU}_{SE} [Table 6]. $\theta^{Acc,max}$, $\theta^{Acc,min}$, and ψ^{Acc} yielded similar values for both BoSE and SE and no appreciable difference was observed [Table 8].

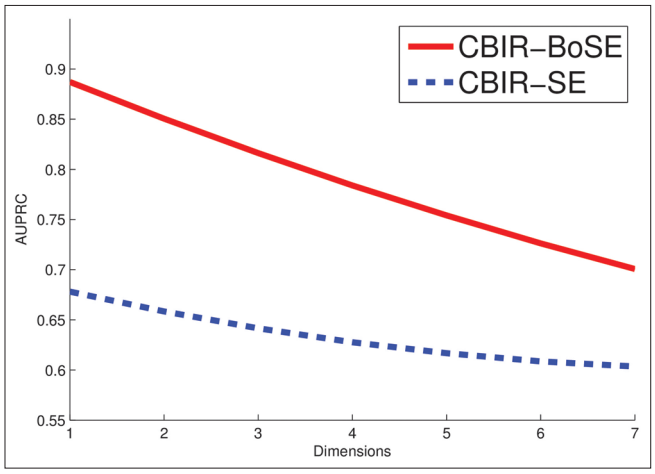

Figure 8: Quantitative results displaying and over the dimensions for Experiment 1. A second order polynomial was fitted to the data to illustrate the trends in

Qualitative evaluation

Figure 12 displays the top five images for both the CBIR-BoSE and CBIR-SE systems. CBIR-BoSE retrieved more relevant images and thus illustrated that images from similar classes are mapped closer to each other in M^{BoSE} compared to M^{SE}. M^{BoSE} [Figure 10b] appears to suggest better class separability compared to SE [Figure 10e].

Experiment 3: Distinguishing High Lymphocytic Infiltration from Low Lymphocytic Infiltration Breast Histopathology

Quantitative evaluation

For D_3, $\theta^{AU,max}_{SE}$ was greater compared to $\theta^{AU,max}_{BoSE}$. $\theta^{AU,min}_{BoSE}$, and ψ^{AU}_{BoSE} were greater compared to $\theta^{AU,min}_{SE}$, and ψ^{AU}_{SE} [Figure 13, Table 6]. $\theta^{Acc,max}_{BoSE}$ and ψ^{Acc}_{BoSE} were higher compared to $\theta^{Acc,max}_{SE}$ and ψ^{Acc}_{SE}, but $\theta^{Acc,min}_{BoSE}$ was similar to $\theta^{Acc,min}_{SE}$ [Table 8]. The dimensionality of the data had little effect on the θ^{Acc}_{BoSE} and θ^{Acc}_{SE}.

Qualitative evaluation

Figure 14 displays the top five images for both the CBIR-BoSE and CBIR-SE systems. M^{BoSE} [Figure 10c] appears to show better separation between the images

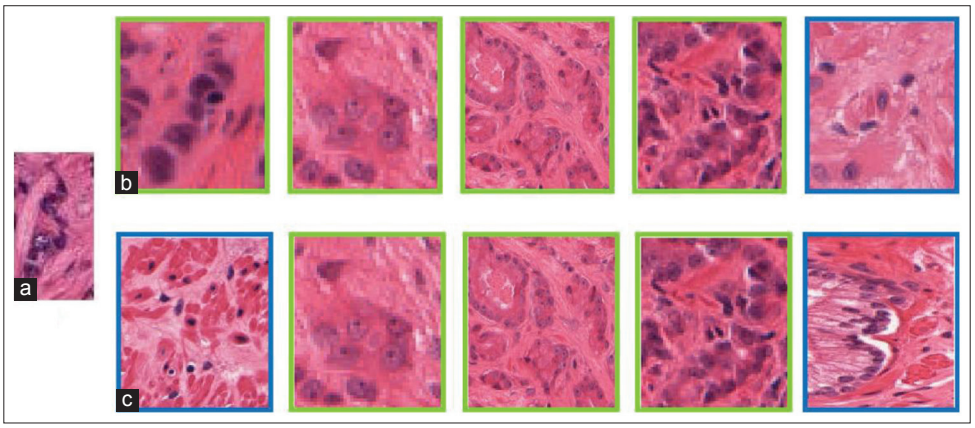

Figure 9:The illustration shows the retrieved images using (b) boosted spectral embedding (BoSE) and (c) spectral embedding (SE) for (a) the query image (prostate cancer tissue sample).The images that are outlined in green and blue are from the cancer and benign classes, respectively. For the top five retrieved images, content-based image retrieval (CBIR)-BoSE returned more relevant images compared to CBIR-SE

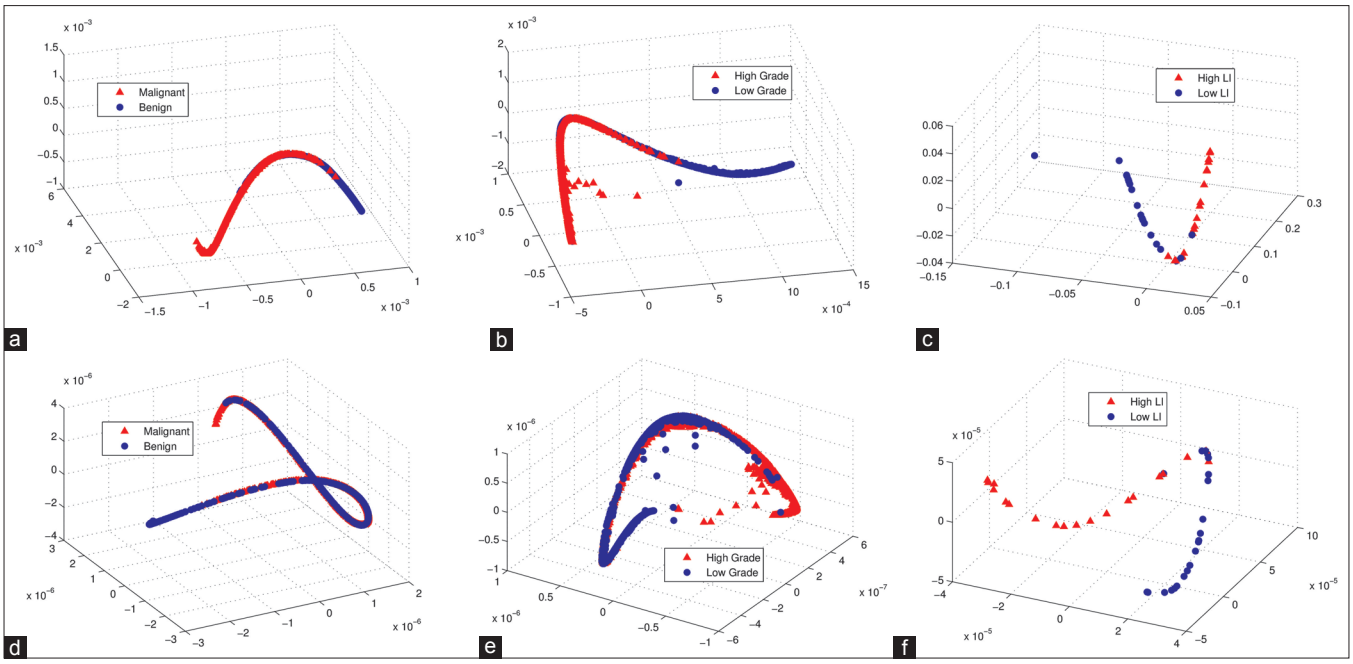

Figure 10: MBoSE and MSE shown for (a), (d) D1, (b), (e) D2, and (c), (f) D3 using (a), (b), (c) BoSE and (d), (e), (f) SE. Although the low-dimensional data do not appear as a set of 'clusters', we can see a clear class separation on the manifold when using BoSE (top row) compared to SE (bottom row). BoSE: Boosted spectral embedding, SE: Spectral embedding

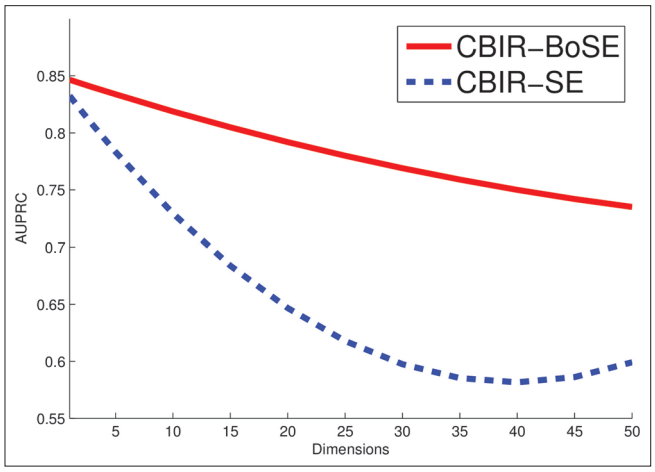

Figure 11: Quantitative results displaying and over all the dimensions for the breast cancer images is greater than. A second order polynomial was fitted to the data to illustrate the trends in. BoSE: Boosted spectral embedding, SE: Spectral embedding

that have low LI and images that have high LI than M^{SE} [Figure 10f].

A Comparison of Content-Based Image Retrieval-BoSE and Content-Based Image Retrieval - Boosted Distance Metric

When comparing CBIR-BoSE against CBIR-BDM, θ^{AU}_{BoSE} and θ^{AU}_{BDM} for D_1 and D_2 were in general comparable to each other with CBIR-BoSE outperforming CBIR-BDM most of the time [Table 7]. However, this comparison is possibly not a fair comparison because the two metrics are being evaluated in different dimensional spaces. Additionally, apart from the marginal outperformance of the BDM via BoSE, it is highly likely that given the high dimensionality of the feature space, CBIR-BDM will most likely also be more unstable compared to CBIR-BoSE.

Area Under Precision-Recall Curves as a Function of Increasing Dimensionality of M^{BoSE}

θ^{AU}_{BoSE} decreased as the dimensionality of M^{BoSE} increased for all three experiments. We offer some intuition as to why this happens. Let the blue triangle in Figure 15 denote the query image. When the dataset is embedded into a one-dimensional space, seven of the eight nearest samples are from the same class. Thus, when performing image retrieval, the majority of the top eight retrieved images will be relevant. When the dataset is embedded into a two-dimensional space, only four of the eight nearest images are from the same class. If image retrieval is performed in this space, only half of the top eight images retrieved will be relevant, reducing precision for that query image; however, classification accuracy for the whole dataset is unchanged. Lastly, when the dataset is embedded into a three-dimensional space, a similar situation is encountered. It should be noted that because classification and training is performed each time a dataset is reduced in dimensionality, it is very possible that all of these spaces will yield either similar classification accuracies or improvements in classification accuracy. Consequently, the apparent discrepancy between the trends in AUPRC and accuracy for BoSE and SE across a different number

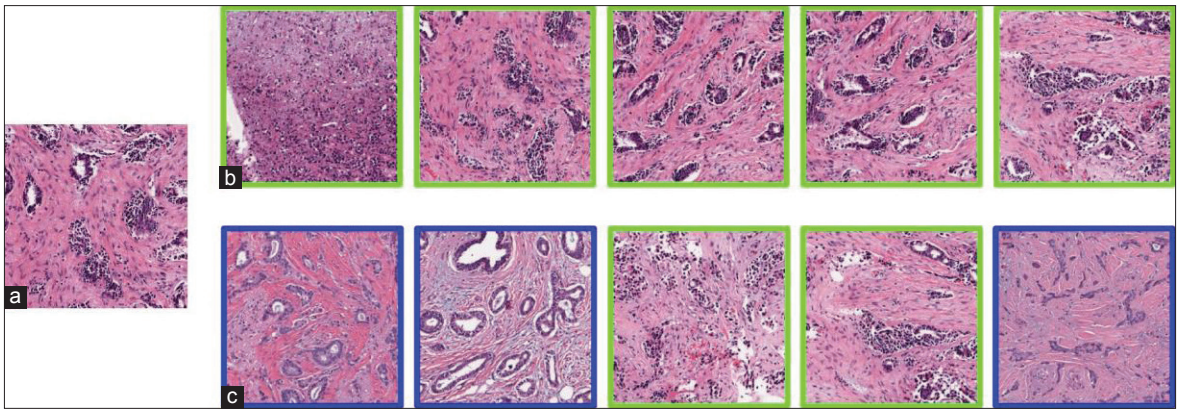

Figure 12: The illustration shows the retrieved images using (b) BoSE and (c) SE for (a) the query image (high grade breast cancer tissue sample). The images that are outlined in green and blue are from high and low grade breast cancer classes, respectively. For the top five retrieved images, CBIR-BoSE returned more relevant images compared to CBIR-SE. BoSE: Boosted spectral embedding, SE: Spectral embedding, CBIR: Content-based image retrieval

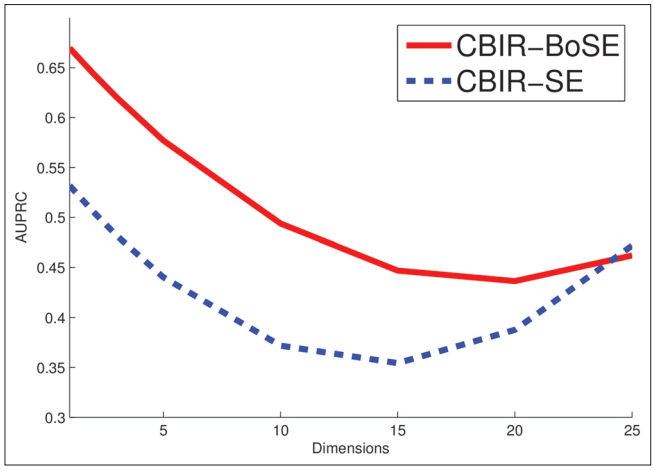

Figure 13: Quantitative results displaying $\theta^{AU}_{BoSE,k}$ and $\theta^{AU}_{SE,k}$ over all the dimensions for the lymphocytic infiltration images. θ^{AU} for BoSE were greater compared to SE. A second order polynomial was fitted to the data to illustrate the trends in θ^{AU}. BoSE: Boosted spectral embedding, SE: Spectral embedding

of dimensions exists because in CBIR the *order* of the retrieved data points affects the AUPRC while the accuracy is unaffected.

CONCLUDING REMARKS

In this paper, we presented a CBIR system that utilized BoSE, which employed the BDM in conjunction with SE. The BDM preferentially weights features that discriminate between objects of different classes allowing for a similarity matrix which better describes object similarity. We have created a task-specific embedding technique that improves class separability, yielding better classification and retrieval. In this work we applied the CBIR-BoSE framework in the context of problems in digital pathology. SE has been shown to be less sensitive to the choice of system parameters compared to other popular manifold learning schemes (e.g. Isomap,[12] LLE[13]). The CBIR system presented here could be employed as a teaching tool for pathology residents and fellows. Specifically, we focused on distinguishing between[1] benign and malignant prostate histology, (2) low and high grade ER + breast cancer histology, and (3) low and high levels of LI in HER2 + breast tissue. We compared CBIR-BoSE to CBIR-SE, which uses the EDM to define object similarity. For different numbers of dimensions of the low-dimensional space, for different

Table 7: Quantitative results showing the maximum, minimum, and mean AUPRC values for Experiment 1 (D_1), Experiment 2 (D_2), and Experiment 3 (D_3). θ^{AU}_{BoSE} is comparable to θ^{AU}_{BDM} for D_1 and D_2. ψ^{AU}_{BoSE} is greater than ψ^{AU}_{BDM} for D_2 and is statistically significant using a $P<0.5$

Dataset	$\theta^{AU,max}_{BoSE}$	$\theta^{AU,max}_{BDM}$	$\theta^{AU,min}_{BoSE}$	$\theta^{AU,min}_{BDM}$	ψ^{AU}_{BoSE}	ψ^{AU}_{BDM}
D_1	0.87	0.87	0.70	0.87	0.79	0.87
D_2	0.90	0.64	0.73	0.64	0.79	0.64
D_3	0.75	0.81	0.45	0.69	0.54	0.77

AUPRC: Area under precision-recall curves

Table 8: Quantitative results showing the maximum, minimum, and mean classification accuracies for Experiment 1 (D_1), Experiment 2 (D_2), and Experiment 3 (D_3). ψ^{Acc}_{BoSE} is greater than ψ^{Acc}_{SE} for D_1 and D_3 and is statistically significant using a $P<0.05$

Dataset	$\theta^{Acc,max}_{BoSE}$	$\theta^{Acc,max}_{SE}$	$\theta^{Acc,min}_{BoSE}$	$\theta^{Acc,min}_{SE}$	ψ^{Acc}_{BoSE}	ψ^{Acc}_{SE}
D_1	0.93	0.81	0.92	0.79	0.93	0.80
D_2	0.99	0.99	0.81	0.76	0.96	0.96
D_3	0.96	0.92	0.90	0.90	0.93	0.91

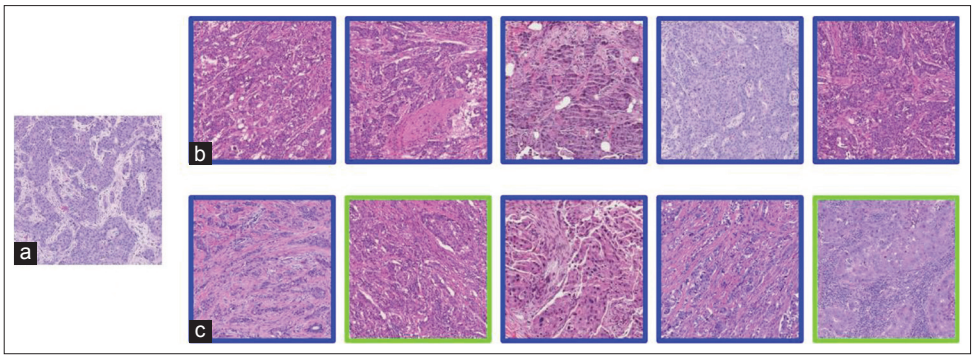

Figure 14: The illustration shows the retrieved images using (b) BoSE and (c) SE for (a) the query image (low LI breast cancer tissue sample). The images that are outlined in green and blue are from the high LI and low LI classes, respectively. In the top five retrieved images, CBIR-BoSE returned more relevant images compared to CBIR-SE. BoSE: Boosted spectral embedding, SE: Spectral embedding, CBIR: Content-based image retrieval

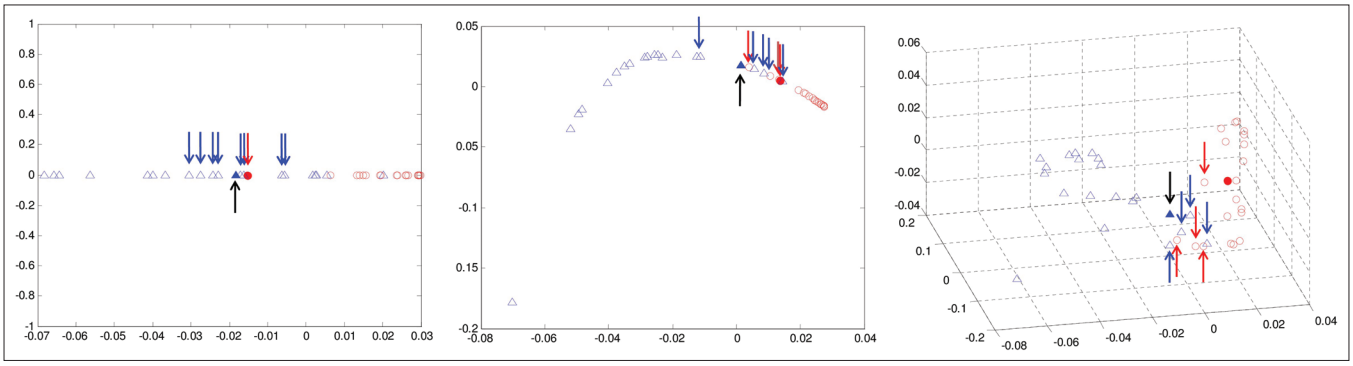

Figure 15: The lymphocytic infiltration data embedded into MBoSE in (a) R1, (b) R2, and (c) R3. The filled in blue triangle denotes the query image and the arrows denote its eight nearest images. When the dimensionality of MBoSE is low, most of the eight nearest images are from the same class as the query image. However, as the dimensions are increased more irrelevant images are part of the query image's eight nearest neighbors. Hence, the area under precision-recall curves decreases as the number of dimensions is increased

datasets, for different performance measures (CBIR and classifier based), CBIR-BoSE outperformed CBIR-SE a majority of the time.

One of the current limitations of our CBIR system is that for every new query image, the manifold for the query along with all existing database images needs to be computed. This procedure needs to be repeated for each new query instance. In future work we are looking to incorporate out of sample extrapolation schemes[40] which allow for the mapping of a new query instance into an existing lower-dimensional space, without having to recompute the eigenvalue decomposition; thus reducing the overall computational cost of a new retrieval task. We also intend to extend our current scheme to the multi-class case.

ACKNOWLEDGEMENTS

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 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. We also wish to thank Dr. John Tomaszewski, Dr. Michael Feldman, Dr. Natalie Shih, Dr. Carolyn Mies, and Dr. Shridar Ganesan for providing and annotating the digitized histopathology data. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

1. Doyle S, Hwang M, Naik S, Feldman M, Tomaszeweski J, Madabhushi A. Using Manifold Learning for Content-based Image Retrieval of Prostate Histopathology. In Workshop on Content-Based Image Retrieval for Biomedical Image Archives (in conjunction with MICCAI); 2007. p. 53-62.
2. Reddy CK, Bhuyan FA. Retrieval and Ranking of Biomedical Images Using Boosted Haar Features. In International Conference on BioInformatics and BioEngineering; 2008. p. 1-6.
3. Yang L, Jin R, Mummert L, Suthankar A, Goode B, Zheng B, *et al.* A boosting framework for visuality-preserving distance metric learning and its application to medical image retrieval. IEEE Trans Pattern Anal Mach Intell 2010;32:30-44.
4. Dy JG, Brodley CE, Kak A, Broderick LS, Aisen AM. Unsupervised feature selection applied to content-based retrieval of lung images. IEEE Trans

Pattern Anal Mach Intell 2003;25:373-8.
5. Bunte K, Petkov N, Biehl M, Jonkman F. Adaptive Metrics for Content Based Image Retrieval in Dermatology. In European Symposium on Aritificial Neural Networks; 2009. p. 129-34.
6. Mehta N, Alomari RS, Chaudhary V. Content based sub-image retrieval system for high resolution pathology images using salient interest points. Eng Med Biol Soc 2009;1:3719-22.
7. Caicedo JC, Gonzalez FA, Romero E. A Semantic Content-based Retrieval Method for Histopathology Images. In Proceedings of the 4th Asia Information Retrieval Conference on Information Retrieval Technology; 2008. p. 51-60.
8. He X, Cai D, Han J. Learning a maximum margin subspace for image retrieval. IEEE Trans Knowl Data Eng 2008;20:189-201.
9. Huang JH, Zia A, Zhou J, Robles-Kelly A. Content-based image retrieval via subspace-projected salient features. Digit Image Comput Tech Appl 2008;593-9.
10. Joliffe I. Principle Component Analysis. New York: Springer-Verlag; 1986.
11. Lee G, Rodriguez C, Madabhushi A. Investigating the efficacy of nonlinear dimensionality reduction schemes in classifying gene and protein expression studies. IEEE Trans Comput Biol Bioinform 2008;5:368-84.
12. Tenenbaum JB, de Silva V, Langford JC. A global geometric framework for nonlinear dimensionality reduction. Science 2000;290:2319-23.
13. Roweis ST, Saul LK. Nonlinear dimensionality reduction by locally linear embedding. Science 2000;290:2323-6.
14. Lafon S, Lee AB. Diffusion maps and coarse-graining: A unified framework for dimensionality reduction, graph partitioning, and data set parameterization. IEEE Trans Pattern Anal Mach Intell 2006;28:1393-403.
15. Shi J, Malik J. Normalized cuts and image segmentation. IEEE Trans Pattern Anal Mach Intell 2000;22:888-905.
16. Ham J, Lee DD, Mika S, Schölkpof B. A kernel view of the dimensionality reduction of manifolds. Int Conf Mach Learn 2004;369-76.
17. Higgs BW, Weller J, Solka JL. Spectral embedding finds meaningful (relevant) structure in image and microarray data. BMC Bioinformatics 2006;7:74.
18. Tiwari P, Kurhanewicz J, Madabhushi A. Multi-kernel graph embedding for detection, Gleason grading of prostate cancer via MRI/MRS. Med Image Anal 2013;17:219-35.
19. ElGhawalby H, Hancock ER. Graph embedding using an edge-based wave Kernel. Int Conf Struct Syntactic Stat Pattern Recogn 2010;6218:60-9.
20. Robles-Kelly A, Hancock ER. A riemannian approach to graph embedding. Pattern Recognit 2007;40:1042-56.
21. Naik J, Doyle S, Basavanhally A, Ganesan S, Feldman M, Tomaszewski JE, *et al.* A boosted distance metric: Application to content based image retrieval and classification of digitized histopathology. SPIE Med Imaging 2009;7260: 72603F1-12.
22. Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. J Comput Syst Sci 1997;55:119-39.
23. 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. Comput Med Imaging Graph 2011;35:506-14.
24. Monaco P, Tomaszewski RE, Feldman MD, Hagemann I, Moradi M, Mousavi P, *et al.* High-throughput detection of prostate cancer in histological sections using probabilistic pairwise markov models. Med Image Anal 2010;14:617-29.
25. Madabhushi A, Doyle S, Lee G, Basavanhally A, Monaco J, Masters S, *et al.* Integrated diagnostics: A conceptual framework with examples. Clin Chem Lab Med 2010;48:989-98.
26. Bloom HJ, Richardson WW. Histological grading and prognosis in breast cancer; a study of 1409 cases of which 359 have been followed for 15 years. Br J Cancer 1957;11:359-77.
27. Kamate C, Baloul S, Grootenboer S, Pessis E, Chevrot A, Tulliez M, *et al.* Inflammation and cancer, the mastocytoma P815 tumor model revisited: Triggering of macrophage activation *in vivo* with pro-tumorigenic consequences. Int J Cancer 2002;100:571-9.
28. Tsuta K, Ishii G, Kim E, Shiono S, Nishiwaki Y, Endoh Y, *et al*. Primary lung adenocarcinoma with massive lymphocytic infiltration. Am J Clin Pathol 2005;123:547-52.
29. Alexe G, Dalgin GS, Scanfeld D, Tamayo P, Mesirov JP, DeLisi C, *et al*. High expression of lymphocyte-associated genes in node-negative her2+ breast cancers correlates with lower recurrence rates. Cancer Res 2007;67:10669-76.
30. Zhang L, Conejo-Garcia JR, Katsaros D, Gimotty PA, Massobrio M, Regnani G, *et al.* Intratumoral T cells, recurrence, and survival in epithelial ovarian cancer. N Engl J Med 2003;348:203-13.
31. Madabhushi A. Digital pathology image analysis: Opportunities and challenges. Imaging Med 2009;1:7-10.
32. Doyle S, Feldman M, Tomaszewski J, Madabhushi A. A boosted Bayesian multiresolution classifier for prostate cancer detection from digitized needle biopsies. IEEE Trans Biomed Eng 2012;59:1205-18.
33. Jain AK, Farrokhnia F. Unsupervised texture segmentation using gabor filters. Pattern Recognit 1991;24:1167-86.
34. Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. IEEE Trans Syst Man Cybern 1973;3:610-21.
35. Basavanhally AN, Ganesan S, Agner S, Monaco JP, Feldman MD, Tomaszewski JE, *et al*. Computerized image-based detection and grading of lymphocytic infiltration in her2+breast cancer histopathology. IEEE Trans Biomed Eng 2010;57:642-53.
36. Hojjatoleslami SA, Kittler J. Region growing: A new approach. IEEE Trans Image Process 1998;7:1079-84.
37. Besag J. On the statistical analysis of dirty pictures. J R Stat Soc 1986;B48:259-302.
38. Doyle S, Feldman MD, Shih N, Tomaszewski J, Madabhushi A. Cascaded discrimination of normal, abnormal, and confounder classes in histopathology: Gleason grading of prostate cancer. BMC Bioinformatics 2012;13:282.
39. Breiman L. Random forests. In: Machine Learning; 2001. p. 5-32.
40. Sparks R, Madabhushi A. Out-of-sample extrapolation using semi-supervised manifold learning (ose-ssl): Content-based image retrieval for prostate histology grading. IEEE Int Symp Biomed Imaging 2011;734-7.