

Symposium - Original Research

Histological stain evaluation for machine learning applications

Jimmy C. Azar, Christer Busch¹, Ingrid B. Carlbom

Department of Information Technology, Centre for Image Analysis, Uppsala University, ¹Department of Pathology and Cytology, University Hospital, Uppsala, Sweden

E-mail: *Jimmy Azar - jimmy.azar@cb.uu.se

*Corresponding author

Received: 23 January 13

Accepted: 23 January 13

Published: 30 March 13

This article may be cited as:

Azar JC, Busch C, Carlbom IB. Histological stain evaluation for machine learning applications. *J Pathol Inform* 2013;4:11.

Available FREE in open access from: <http://www.jpathinformatics.org/text.asp?2013/4/2/11/109869>

Copyright: © 2013 Azar JC. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Aims: A methodology for quantitative comparison of histological stains based on their classification and clustering performance, which may facilitate the choice of histological stains for automatic pattern and image analysis. **Background:** Machine learning and image analysis are becoming increasingly important in pathology applications for automatic analysis of histological tissue samples. Pathologists rely on multiple, contrasting stains to analyze tissue samples, but histological stains are developed for visual analysis and are not always ideal for automatic analysis. **Materials and Methods:** Thirteen different histological stains were used to stain adjacent prostate tissue sections from radical prostatectomies. We evaluate the stains for both supervised and unsupervised classification of stain/tissue combinations. For supervised classification we measure the error rate of nonlinear support vector machines, and for unsupervised classification we use the Rand index and the F-measure to assess the clustering results of a Gaussian mixture model based on expectation-maximization. Finally, we investigate class separability measures based on scatter criteria. **Results:** A methodology for quantitative evaluation of histological stains in terms of their classification and clustering efficacy that aims at improving segmentation and color decomposition. We demonstrate that for a specific tissue type, certain stains perform consistently better than others according to objective error criteria. **Conclusions:** The choice of histological stain for automatic analysis must be based on its classification and clustering performance, which are indicators of the performance of automatic segmentation of tissue into morphological components, which in turn may be the basis for diagnosis.

Key words: Support vector machines, expectation-maximization, Gaussian mixture model, F-measure, Rand index, Mahalanobis distance, Fisher criterion, high throughput imaging systems

Access this article online

Website:
www.jpathinformatics.org

DOI: 10.4103/2153-3539.109869

Quick Response Code:



INTRODUCTION

Pathologists rely on histological stains to identify morphological changes in tissue that are linked to cancer and other diseases. Different stain combinations target different types of tissue. With the most common stain,

hematoxylin/eosin, the hematoxylin stains the cell nuclei blue, and the counter-stain, eosin, stains the cytoplasm pink and stromal components in various grades of red/pink. This and most other stains are developed for visual examination of the tissue under a microscope, and thus they are often not optimal for automated

pattern and image analysis. However automated and computer-assisted analysis of histopathology specimens is gaining importance since the introduction of high throughput imaging systems.

As a first step in the automated analysis of tissue samples, morphological components, such as nuclei, stroma, and cytoplasm, often need to be identified and segmented from the surrounding tissue for further analysis. This identification relies on the absorption of histological stains by the tissue components, and therefore the choice of stain has a significant effect on the outcome of the segmentation. Color image analysis for histologically stained tissue is further complicated by both inter- and intra-specimen intensity variations due to variations in the tissue preparation process and overlap in the stain absorption spectra.^[1,2]

Quantitative evaluation of the efficacy of a particular stain for automatic color image analysis relies upon the evaluation of the stained tissue in a color space where the distance between two points represents the chromaticity difference between the corresponding colors, reducing the dependence on the intensity variations that may arise from tissue preparation and image acquisition. In such a color space, the tissue image data forms distinct clusters, one cluster for each stain/tissue combination in the image. With an ideal stain, these clusters are compact and distant from each other, enabling a machine learning algorithm to separate the clusters and yield an accurate color segmentation.

We explore a methodology for selecting optimal histological stains for automation in terms of both supervised and unsupervised classification performances. We use three types of quantitative efficacy measures: The classification error for supervised classification, the F-measure and the Rand index for clustering performance, and the Fisher criterion and Mahalanobis distance for cluster separability. We demonstrate these measures on 13 different stains applied to prostate tissue. Although these stains are not all typically used for prostate cancer malignancy grading, they are included for illustration purposes. We conclude that regardless of

the classification method, the same stains perform the best for this type of tissue; for other types of tissues a different set of stains would undoubtedly perform better.

There are numerous methods in the literature dealing with color decomposition, or deconvolution.^[1-5] However they do not attempt to compare the different stains from a classification perspective. Our aim is to demonstrate a methodology for selecting an optimal stain for color decomposition. To our knowledge, there has not been a similar attempt at comparing histological stains from a machine learning point of view.

MATERIALS AND METHODS

Our data comprises adjacent sections of prostate tissue from radical prostatectomies prepared by the Department of Pathology, University Hospital, Uppsala, Sweden. The sections were stained with 13 different histological stains and scanned using the Aperio ScanScope XT Slide Scanner (Aperio Technologies, Vista, CA) with a $\times 40$ objective. The histological stains were hematoxylin/eosin (H&E), Mallory's trichrome (MALLORY), Miller's elastic (MILLERS-E), Van Gieson (VG), Van Gieson elastic (VG-E), Giemsa-eosin stain (GIEMSA), Grocott's methenamine silver stain (GROCOTT), hematoxylin-Herovici polychrome (HERO), Weigert's elastic (WEIGERT), Cytokeratin immunohistochemistry (CYTK), Periodic acid-Schiff (PAS), Ladewig (LADEWIG), and Sirius-hematoxylin (SIR+), an experimental stain combination we have tested. For information on histological stains, the reader is referred to Bancroft *et al.*^[6] Images of tissue samples stained with hematoxylin-eosin, Mallory's trichrome, and Sirius-hematoxylin are shown in Figure 1.

Our methodology for comparing different stains is based on assessing the performance of both supervised and unsupervised classifications on the given dataset. In our case, we are interested in three stain/tissue combinations, or classes, namely nuclei, stroma, and cytoplasm. For the supervised case, we require a training set constituting

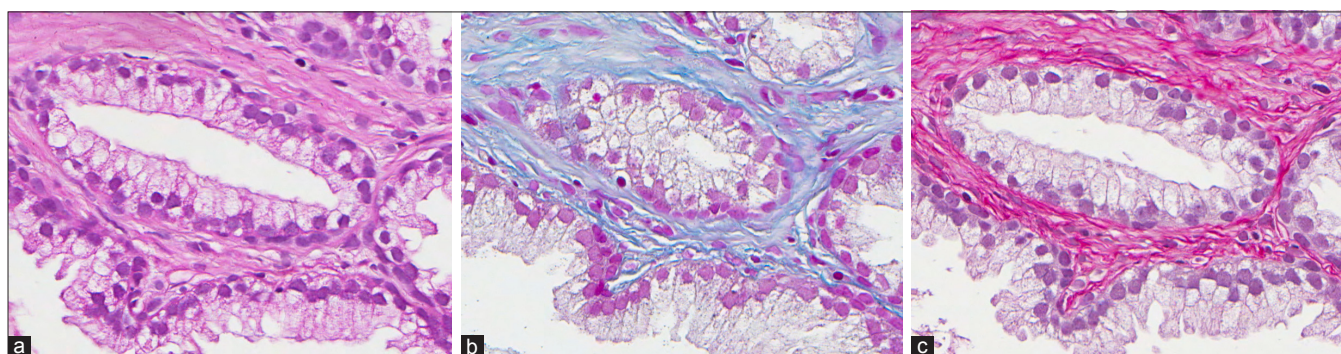


Figure 1: (a) Hematoxylin-eosin stain. (b) Mallory's trichrome stain. (c) Sirius-hematoxylin stain.

the ground truth as seen by a pathologist, and this was obtained by the manual selection of 30 different regions in each of the stained 640×816 pixel images and by the labeling of each region according to its class. The regions were selected completely within the target, with 1000 pixels selected per class. We train and optimize a support vector classifier over the data from each stained image and report the classification error using ten-fold cross-validation. We chose the support vector classifier with a radial basis function, as it provides a significant range of complexity that may be controlled by optimizing the kernel and regularization parameters based on the cross-validated classification error. This allows the classifier to generalize well and avoid overfitting. For the unsupervised case, we assess the clustering performance of the Gaussian mixture model by comparing the cluster labels with those of the ground truth by using precision and recall. We use the Gaussian mixture model, as it is not constrained by the assumption of spherical clusters as are the k-means and fuzzy c-means. For cluster separability we use two measures, the Fisher Criterion and the Mahalanobis distance.

RESULTS

Feature Space

To eliminate dependency on inter- and intra-stain variations, which is essential for robust classification, we adopt the Maxwellian plane as our feature space. We first transform the image RGB data to a linear model using the Beer-Lambert Law.^[7] Next we transform the resulting color data to the Maxwellian plane^[8] at a distance of $1/\sqrt{3}$ from the origin, using the perspective transformation centered at the origin.^[9,2]

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & \frac{\sqrt{2}}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}.$$

This decouples the intensity from the color information, thereby eliminating the dependence on intensity. As a result, the distance between two points in the Maxwellian plane corresponds to the chromaticity difference between the corresponding colors. Figure 2a shows the scatter plots for the color data in a section stained with MILLERS-E in the RGB space, the Beer-Lambert space, and the Maxwellian chromaticity plane. It is evident that there is considerable overlap between the color regions corresponding to the cytoplasm and the nuclei in the Maxwellian plane. This intrinsic class overlap limits the performance of the classification. In Figure 2b, the corresponding scatter plots for CYTK show that the classes are well separated in the Maxwellian plane and consequently the classification should yield a much better result.

Supervised Classification

To assess classification performance, a support vector machine^[10,11] was utilized for each stain. Given an image with a particular stain, let $\{\mathbf{x}_i, i = 1, \dots, n\}$ denote the set of vectors representing the pixel locations in the image (i.e., patterns) where n is the total number of pixels. In the Maxwellian plane, each such vector \mathbf{x}_i is in \mathbb{R}^2 . For every pair of classes, ω_1, ω_2 , a general support vector classifier is defined as $f(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) + w_0$, where \mathbf{x} is a pattern to be classified, \mathbf{w}, w_0 are the weights defining the classifier, and $\phi(\cdot)$ is a mapping of the patterns into kernel space. We have used a radial basis function (RBF) defined by:

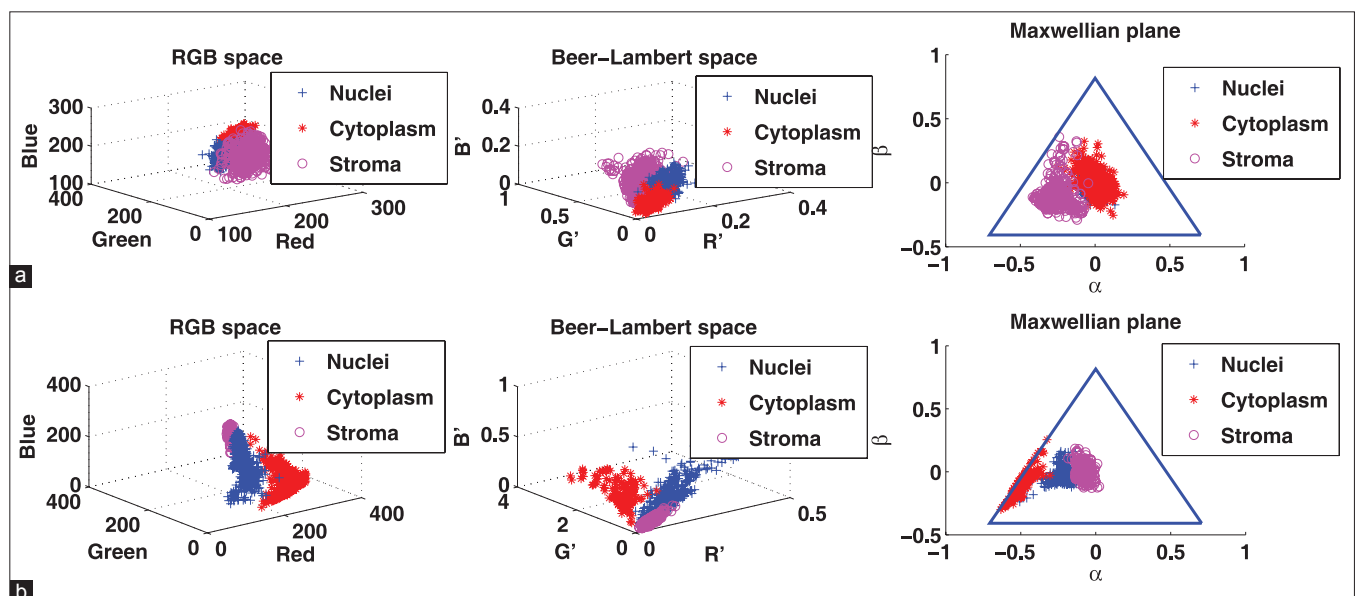


Figure 2: (a-b) Scatter plots in RGB space (left), Beer-Lambert space (center), and Maxwellian plane (right). (a) MILLERS-E. (b) CYTK

$$K(\mathbf{x}_i, \mathbf{x}) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}\|^2)$$

for the support vector classifier, where γ is a parameter that controls the width of the Gaussian kernel. The classifier may be optimized with regard to γ using ten-fold cross-validation; however, the classifier parameters in this case consist not only of γ , but also of a regularization term C , which may be chosen freely (please refer to the Appendix for details). In order to optimize the classifier with regard to both these parameters simultaneously, an exhaustive grid search algorithm^[12] ensures arriving at the optimal combination of the (C, γ) values. An example of such an optimization is shown in Figure 3 in the form of a contour plot in which the minimum corresponds to the optimal combination.

For each stain, we construct a support vector classifier using an RBF kernel that is then optimized as described above, with γ ranging from 2^{-15} to 2^5 and the regularization parameter C ranging from 2^{-5} till 2^9 , to ensure coverage of a wide range of simple-to-complex boundaries.^[12] We then perform ten-fold cross-validation on each stain using a support vector classifier optimized for that stain. The final result is shown in Figure 4, where the abscissa is ordered according to increasing classification error. We conclude that with respect to supervised classification, the stains MALLORY, CYTK, HERO, GIEMSA, and SIR+ show a considerably better performance for prostate tissue than the other stains. This is attributed to a significant class overlap for the other stains. A stain combination that stains several different structures in the tissue with the same color component is at a disadvantage, as this may lead to a high class overlap and low classification rates.

Unsupervised Classification

Blind classification is sometimes preferred over supervised classification because it avoids manual labeling. Assuming that the class labels are unknown, we assess the clustering performance by the Gaussian mixture model.^[13,14] For each class, ω_j , the probability density function is modeled by a Gaussian component completely described by its mean μ_j and covariance matrix Σ_j :

$$p(\mathbf{x}; \theta_j) = \frac{1}{\sqrt{(2\pi)^d |\Sigma_j|}} e^{-\frac{1}{2}(\mathbf{x}-\mu_j)^T \Sigma_j^{-1}(\mathbf{x}-\mu_j)},$$

where θ_j is the set of parameters for the distribution of class ω_j , that is, $\theta_j = \{\mu_j, \Sigma_j\}$, and d is the number of features, that is, the α and β coordinates in the Maxwellian plane. Thus, the complete dataset may be described as a mixture of these components given by: $p(\mathbf{x}; \Psi) = \sum_{j=1}^K \alpha_j \cdot p(\mathbf{x}; \theta_j)$ where K is the number of Gaussian components, or clusters, Ψ is the total set of parameters for the mixture distribution, and α_j is the prior probability associated with the j^{th} Gaussian component, and $\sum_{j=1}^K \alpha_j = 1$. The expectation-maximization algorithm is used to maximize the log-likelihood, Q , by observing the data given by the Gaussian model. During the expectation step (or E-step) of the algorithm, the cluster membership of each object, w_{ji} , is computed using the Bayes' theorem. This is followed by an update of the mixture prior probabilities, α_j , and the cluster means and covariance matrices, $\theta_j = \{\mu_j, \Sigma_j\}_{j=1, \dots, K}$ during the maximization step (or M-step). The E-M steps are repeated in succession until convergence, that is, until the change in the log-likelihood function Q is below a tolerance of 10^{-10} (see Appendix for details). We initialize the expectation-maximization algorithm with repeated random initializations to obtain the first estimates for the mixture density parameters and select the result with the highest log-likelihood. The results of the Gaussian mixture model clustering for the MALLORY, GIEMSA, SIR+, and CYTK stains, on the prostate tissue samples in Figure 1, are shown in Figure 5. The highlighted contours in the figures correspond to a level 25% below the maximum value of the Gaussian.

To assess clustering performance objectively, we use the class labels from the previous supervised classifications and compute two cluster validation measures, the Rand index^[15] and the F-measure^[16,17], which are computed pairwise over all the objects in the dataset:

$$\text{Rand Index: } RI = \frac{TP + TN}{TP + FP + FN + TN},$$

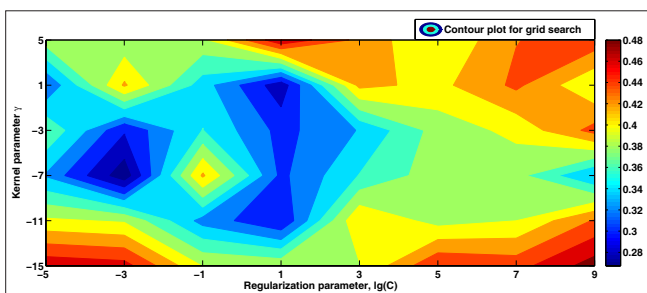


Figure 3: Optimization of SVM radial basis function kernel parameter and regularization parameter C by grid search. Note that $\lg(\cdot)$ is the base 2 logarithm

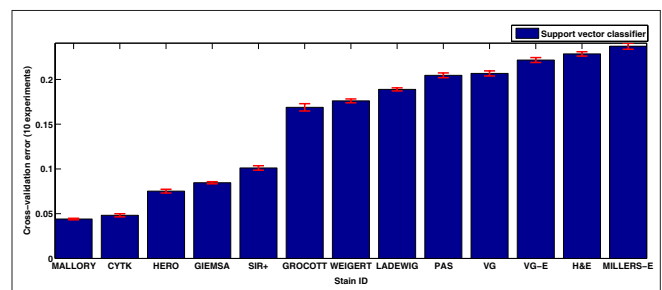


Figure 4: Ten-fold cross-validation error using an optimized SVM classifier for each stain

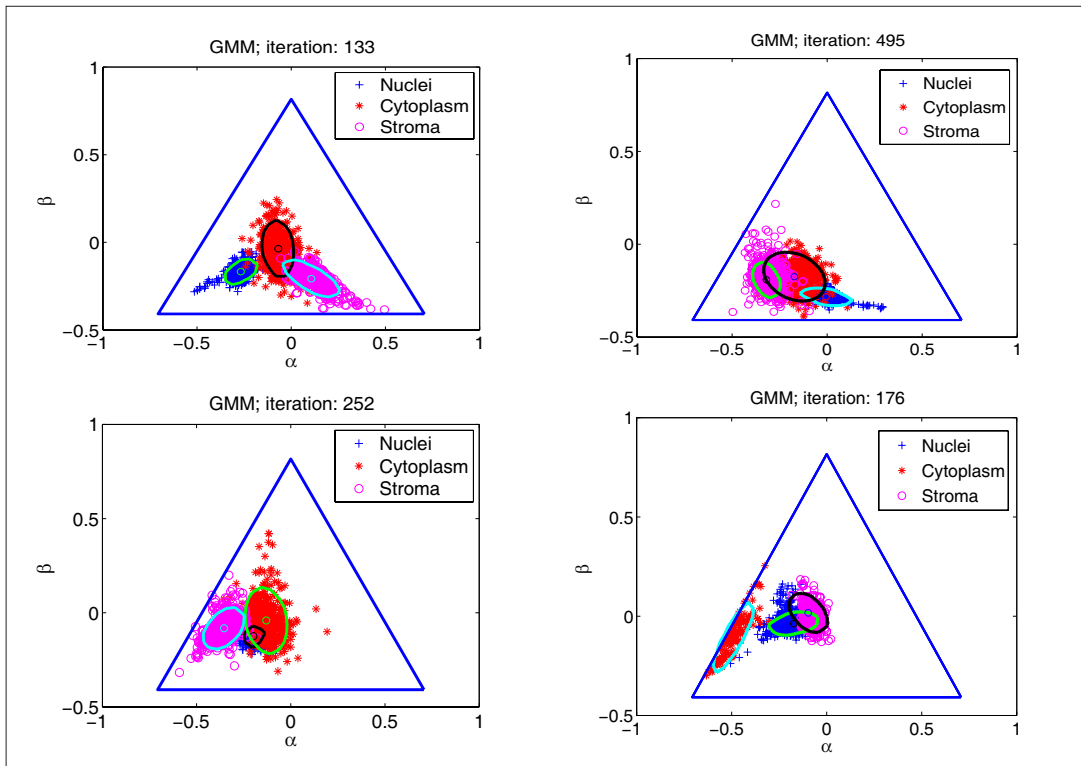


Figure 5: Gaussian mixture model clustering for different stains: Mallory (upper left), Giemsa (upper right), SIR+ (lower left), CYTK (lower right)

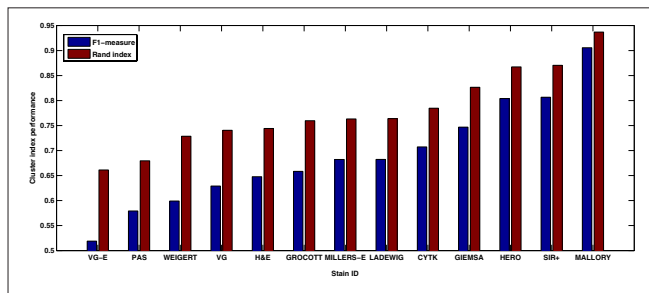


Figure 6: Clustering evaluation using the F_1 -measure and Rand index

where $TP = True\ Positives$, $TP = True\ Negatives$, $FP = False\ Positives$, and $FN = False\ Negatives$, and

$$F\text{-measure: } F_{\beta} = \frac{(1 + \beta^2) \cdot P \cdot R}{\beta^2 P + R},$$

where P and R are precision and recall, defined as:

$$Precision: P = \frac{TP}{TP + FP}, \text{ and}$$

$$Recall: R = \frac{TP}{TP + FN}.$$

We chose $\beta = 1$ to give comparable weights to precision and recall, and the measure is then denoted as an F_1 -measure. The result of this analysis is shown in Figure 6. Both performance measures indicate that

the five best stains in the supervised case are also more favorable here.

Class Separability Measures

In order to assess the separability of the classes in terms of how compact and distant they are from each other, we have utilized a Fisher-based criterion^[18] and a sum of the squared Mahalanobis distance.^[19] We define the Fisher criterion as $\text{trace}(\mathbf{S}_W^{-1}\mathbf{S}_B)$ where \mathbf{S}_W and \mathbf{S}_B are the within- and between-scatter matrices:

$$\mathbf{S}_W = \sum_{i=1}^K \frac{n_i}{n} \mathbf{\Sigma}_i, \quad \text{and}$$

$$\mathbf{S}_B = \sum_{i=1}^K \frac{n_i}{n} (\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T,$$

with $\boldsymbol{\mu}$ the mean of the entire dataset, and $\mathbf{\Sigma}_i$ and $\boldsymbol{\mu}_i$, the covariance matrix and mean of class ω_i , respectively, n_i the number of patterns in class ω_i , and n the total number of patterns in the dataset. On the other hand, the Mahalanobis distance defined between two classes with means $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$ and covariance matrices $\mathbf{\Sigma}_1, \mathbf{\Sigma}_2$ is given by:

$$D_M = (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^T \left(\frac{\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2}{2} \right)^{-1} (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1).$$

If the classes are compact and well-separated, then they should have a relatively high Mahalanobis distance. We compute the squared Mahalanobis distance for every

pair of classes and sum up these values. The result is shown in Table 1, where we see that the optimal stains according to these measures are Mallory's trichrome and Cytokeratin. However, we also see that some of the stains, such as the Miller's elastic, rank well despite their poor classification rate, and this is attributed to the fact that the scatter-based criteria may not always predict the performance of a classifier correctly. A simple example that demonstrates this occurs when two classes are compact, but completely overlapping, and the third class is compact and extremely distant from the other two classes. This third class would influence the result positively, although two of the classes are completely overlapping, and will lead to a low classification rate. One can notice the effect a distant class may have on the Mahalanobis distance by observing the scatter plot of the Cytokeratin stain in the Maxwellian plane [Figure 5], which has led to the high value of the Mahalanobis distance for CYTK in Table 1.

DISCUSSION

We have presented a methodology for quantitative comparison of histological stains based on their classification and clustering performance. Among the stains we have investigated, using prostate tissue samples, the classification results suggest that MALLORY, CYTK, HERO, GIEMSA, and SIR+ have an advantage over the other stains, with supervised classification error rates below 11%, and below 5% for MALLORY and CYTK. Furthermore, blind classification was assessed using the Gaussian mixture model and the Rand and F_1 -measures.

Table 1: Values for the fisher criterion and sum-of-squared mahalanobis distance for the different stains

Stain	Maxwellian plane	
	Fisher criterion	Mahalanobis distance
Mallory	7.7079	70.1254
CYTK	6.6916	108.9913
HERO	6.4888	58.4788
MILLERS-E	5.0686	45.6796
GROCOTT	4.5433	40.9382
GIEMSA	4.2874	38.9782
SIR+	3.2531	29.3748
WEIGERT	3.0074	27.2329
PAS	1.4119	12.7243
VG	1.3252	11.9435
VG-E	1.2972	11.7466
LADEWIG	1.0858	9.7851
H&E	1.0351	9.3282

CYTK: Cytokeratin immunohistochemistry, HERO: hematoxylin-Herovici polychrome, MILLERS-E: Miller's elastic, GROCOTT: Grocott's methenamine silver stain, GIEMSA: Giemsa-eosin stain, SIR+: Sirius-hematoxylin, WEIGERT: Weigert's elastic, PAS: Periodic acid-Schiff, VG: Van Gieson, LADEWIG: Ladewig, H&E: Hematoxylin-eosin

Results show that MALLORY and SIR+ outperform other stains in terms of clustering accuracy, with Rand indices above 0.85. We also saw that simple criteria, such as the Fisher criterion and the Mahalanobis distance, give a good ranking of stains, but may in some cases be misleading. The ideal stain is one where the reference colors are compact and separate, with a good classification performance. Furthermore, with an ideal stain, all tissue types absorb a similar amount of stain, without oversaturation. For the purpose of high-throughput imaging, the choice of a histological stain for a given application must be motivated by the objective criteria that optimize the automation process by facilitating machine learning and image analysis techniques.

ACKNOWLEDGMENT

The authors wish to thank Ulrika A. Larsson at the Uppsala University Hospital for preparing the tissue samples. This research was supported by Vetenskapsrådet under grant 2009-5418.

APPENDIX

Support Vector Machines

The correct classification of all objects is assumed if the following conditions hold for $i = 1, \dots, n$:

$$\begin{cases} \mathbf{w}^T \phi(\mathbf{x}) + w_0 \geq 1 - \xi_i, & y_i = +1 \\ \mathbf{w}^T \phi(\mathbf{x}) + w_0 \leq -1 + \xi_i, & y_i = -1 \\ \xi_i \geq 0 \end{cases}$$

where y_i is the class label for object \mathbf{x}_i , ξ_i and is the slack variable corresponding to object \mathbf{x}_i . These conditions may be rewritten in compact form as

$$y_i(\mathbf{w}^T \phi(\mathbf{x}) + w_0) - 1 + \xi_i \geq 0, \quad i = 1, \dots, n.$$

The support vector classifier aims at maximizing the margin between the classes, which is equivalent to minimizing the norm of the weights. In addition, the slack variables are added to the quantity to be minimized, while regulated by a cost (or regularization) parameter C to allow for class overlap. This minimization is subject to constraints, namely, the patterns are classified correctly, as described above. Thus the classifier minimizes the following criterion:

$$\text{minimize } \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i,$$

$$\text{subject to } y_i(\mathbf{w}^T \phi(\mathbf{x}) + w_0) - 1 + \xi_i \geq 0, \quad i = 1, \dots, n.$$

This may be formulated using Lagrange multipliers, whereby the following Lagrangian form is obtained:

$$L = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i (y_i (\mathbf{w}^T \phi(\mathbf{x}_i) + w_0) - 1 + \xi_i) - \sum_{i=1}^n r_i \xi_i,$$

where α_i is the Lagrange multiplier associated with the constraint equation corresponding to pattern \mathbf{x}_i , and r_i is the Lagrange multiplier associated with slack variable ξ_i and ensures that ξ_i remains positive. By differentiating L with respect to the weights and the slack variables, setting the result to zero, and then solving, we obtain the following:

$$\begin{cases} \frac{\partial L}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{w} = \sum_{i=1}^n \alpha_i y_i \phi(\mathbf{x}_i) \\ \frac{\partial L}{\partial w_0} = 0 \Rightarrow \sum_{i=1}^n \alpha_i y_i = 0 \\ \frac{\partial L}{\partial \xi_i} = 0 \Rightarrow C - \alpha_i - r_i = 0 \end{cases} .$$

Note that the last result $C - \alpha_i - r_i = 0$ may be combined with $r_i \geq 0$ to give $0 \leq \alpha_i \leq C$. By substituting these results back into the Lagrangian, we obtain the dual form of the Lagrangian:

$$L_D = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j).$$

Quadratic programming is then employed to solve for the Lagrange multipliers, α_i , $i = 1, \dots, n$. Once these are obtained, \mathbf{w} is computed from $\mathbf{w} = \sum_{i \in SV} \alpha_i y_i \phi(\mathbf{x}_i)$, where SV is the set of support vectors (objects associated with $\alpha_i \neq 0$). Thereafter, the offset w_0 is obtained from: $\alpha_i (y_i (\mathbf{w}^T \phi(\mathbf{x}_i) + w_0) - 1 + \xi_i)$ by averaging over objects with $0 < \alpha_i < C$, that is, $\xi_i = 0$. The support vector classifier is then derived as:

$$f(\mathbf{x}) = \sum_{i \in SV} \alpha_i y_i \phi^T(\mathbf{x}_i) \phi(\mathbf{x}) + w_0.$$

The scalar product $\phi^T(\mathbf{x}_i) \phi(\mathbf{x})$ is substituted by a kernel function, $K(\mathbf{x}_i, \mathbf{x}) = \phi^T(\mathbf{x}_i) \phi(\mathbf{x})$ without the need to explicitly define the transformation function or mapping $\phi(\cdot)$.

Gaussian Mixture Model

The likelihood of observing the data given by the model for a single component is given by:

$$L(\mathbf{x}_1, \dots, \mathbf{x}_n) = p(\mathbf{x}_1, \dots, \mathbf{x}_n; \psi) = \prod_{i=1}^n p(\mathbf{x}_i; \psi) = \prod_{i=1}^n \sum_{j=1}^K \alpha_j \cdot p(\mathbf{x}_i; \theta_j).$$

If we assume that the class labels are known and are given by:

$$z_{ji} = p(\omega_j | \mathbf{x}_i) = \begin{cases} 1, & \text{if } \mathbf{x}_i \in \omega_j \\ 0, & \text{otherwise,} \end{cases}$$

then the likelihood function may be rewritten as:

$$L(\mathbf{x}_1, \dots, \mathbf{x}_n) = \prod_{i=1}^n \prod_{j=1}^K [\alpha_j \cdot p(\mathbf{x}_i; \theta_j)]^{z_{ji}}.$$

The Gaussian mixture model employs the expectation-maximization algorithm to maximize this likelihood or rather log-likelihood function

$$LL(\mathbf{x}_1, \dots, \mathbf{x}_n) = \sum_{i=1}^n \sum_{j=1}^K z_{ji} \cdot \log(\alpha_j) + \sum_{i=1}^n \sum_{j=1}^K z_{ji} \cdot p(\mathbf{x}_i; \theta_j).$$

However, since z_{ji} is not known, the following function is maximized instead:

$$Q = \sum_{i=1}^n \sum_{j=1}^K w_{ji} \cdot \log(\alpha_j) + \sum_{i=1}^n \sum_{j=1}^K w_{ji} \cdot p(\mathbf{x}_i; \theta_j).$$

During the E-step, the cluster membership of each object, w_{ji} at any given iteration m becomes:

$$w_{ji} = p(\omega_j | \mathbf{x}_i, \psi^m) = \frac{\alpha_j^{(m)} \cdot p(\mathbf{x}_i | \omega_j; \theta_j^{(m)})}{\sum_{k=1}^K \alpha_k^{(m)} \cdot p(\mathbf{x}_i | \omega_k; \theta_k^{(m)})}.$$

During the M-step of the algorithm the distribution parameters are updated:

$$\begin{aligned} \alpha_j &= \frac{1}{n} \sum_{i=1}^n w_{ji}, \\ \boldsymbol{\mu}_j &= \frac{\sum_{i=1}^n w_{ji} \cdot \mathbf{x}_i}{\sum_{i=1}^n w_{ji}}, \text{ and} \\ \boldsymbol{\Sigma}_j &= \frac{\sum_{i=1}^n w_{ji} \cdot (\mathbf{x}_i - \boldsymbol{\mu}_j)(\mathbf{x}_i - \boldsymbol{\mu}_j)^T}{\sum_{i=1}^n w_{ji}}. \end{aligned}$$

These two steps are repeated iteratively until convergence.

REFERENCES

1. Begelman G, Zibulevsky M, Rivlin E, Kolatt T. Blind decomposition of transmission light microscopic hyperspectral cube using sparse representation. *IEEE Trans Med Imag* 2009;28:1317-24.
2. Gavrilovic M. Spectral image processing with applications in biotechnology and pathology. Doctoral thesis. Uppsala:Acta Universitatis Upsaliensis; 2011.
3. Tadrous PJ. Digital stain separation for histological images. *J Microsc* 2010;240:164-72.
4. Rabinovich A, Laris CA, Inc Qdm, Agarwal S, Price JH, Belongie S. Unsupervised color decomposition of histologically stained tissue samples. *Advances in Neural Information Processing Systems*; 2004.
5. Ruifrok AC, Johnston DA. Quantification of histochemical staining by color deconvolution. *Anal Quant Cytol Histol* 2001;23:291-9.
6. Bancroft JD, Gamble M. Theory and practice of histological techniques. 6th ed. Churchill Livingstone: Elsevier; 2007.
7. Parson WW. Modern optical spectroscopy. New York: Springer; 2009.
8. Judd DE. A maxwell triangle yielding uniform chromaticity scales. *J Opt Soc Am* 1935;25:24-35.
9. Foley J, van Dam A, Feiner S, Hughes J. Computer graphics: Principles and practice. Massachusetts: Addison Wesley; 1992. p. 213.
10. Vapnik V. The Nature of statistical learning theory. New York: Springer-Verlag;

- 1995.
11. Bishop C. Pattern recognition and machine learning. New York: Springer; 2006. p. 325.
12. Hsu CW, Chang CC, Lin CJ. A practical guide to support vector classification. Tech report. Dept of CS and IE, National Taiwan University, Taipei. 2003. Available from: <http://www.csie.ntu.edu.tw/~cjlin/libsvm/> [Last accessed on 2012 Sep 4].
13. Theodoridis S, Koutroumbas K. Pattern Recognition. 4th ed. Massachusetts: Academic Press; 2009. p. 44-8.
14. Bishop C. Statistical pattern recognition. New York: Springer; 2006. p. 430.
15. Rand WM. Objective criteria for the evaluation of clustering methods. *J Am Stat Assoc* 1971;66:846-50.
16. van Rijsbergen CJ. Information retrieval. 2nd ed. London: Butterworths; 1979.
17. Curic V, Heilio M, Krejic N, Nedeljkov M. Mathematical Model for efficient water flow management. *Nonlinear Anal Real World Appl* 2010;11:1600-12.
18. Webb A. Statistical pattern recognition. 2nd ed. USA: Wiley; 2002. p. 311.
19. Webb A. Statistical pattern recognition, 2nd ed. USA: Wiley; 2002. p. 414.