Histology Image Classification using Supervised Classification and Multimodal Fusion

Tao Meng, Lin Lin, Mei-Ling Shyu
Department of Electrical and Computer Engineering
University of Miami
Coral Gables, FL 33124, USA
Email: {t.meng, l.lin2}@umiami.edu, shyu@miami.edu

Shu-Ching Chen
School of Computing and Information Sciences
Florida International University
Miami, FL 33199, USA
Email: chens@cs.fiu.edu

***Abstract*—The fast development of microscopy imaging techniques nowadays promotes the generation of a large amount of data. These data are very crucial not only for theoretical biomedical research but also for clinical usage. In order to decrease the inter-intra observer variability and save the human effort on labeling and classifying these images, a lot of research efforts have been devoted to the development of algorithms for biomedical images. Among such efforts, histology image classification is one of the most important areas due to its broad applications in pathological diagnosis such as cancer diagnosis. To improve classification accuracy, most of the previous work focuses on extracting more features and building algorithms for a specific task. This paper proposes a framework based on the novel and robust Collateral Representative Subspace Projection Modeling (C-RSPM) supervised classification model for general histology image classification. In the proposed framework, a cell image is first divided into 25 blocks to reduce the spatial complexity of computation, and one C-RSPM model is built on each block set which contains blocks in the same location from different images. For each testing image, our proposed framework first classifies each of its blocks using the C-RSPM classification model built for that block set, and then applies a multimodal late fusion algorithm with a weighted majority voting strategy to decide the final class label of the whole image. Experimenting using three-fold cross validation with three benchmark histology data sets shows that the proposed framework outperforms other well-known classifiers in the comparison and gives better results than the highest accuracy reported previously.**

Keywords-histology image classification; multimodal fusion; weighted majority voting algorithm; C-RSPM

I. INTRODUTION

With the development of automated imaging technologies, such as the fluorescent microscopy, two-photon-laser scanning microscopy and electron microscopy, there is a rapid growth in imaging data on a scale comparable to that of the genomic revolution [1]. To cope with the enormous amount of bio image data, the computer based image analysis is required to improve the working efficiency and to decrease inter-reader variability. Therefore, quite a lot of software and applications have been developed for different tasks like cell detection [2], [3], bio-image segmentation [4], [5] and cell phenotype classification [6], [7]. Recently, the collaboration between biomedical researchers and computer engineers has pushed the development of the area of Bioimage informatics forward and developed many useful applications. A broad review is given in [8].

Histology is the science of understanding the structure of biological organisms and studying functional semantics of different structures by examining the microscopic anatomy of cells and tissues. In histology analysis, the bio tissue is usually prepared as thin slices and goes through a staining process, such as Hematoxylin and Eosin (H&E) staining [9] for better imaging under microscope. Since significant changes of functions of any kind are manifested with changes in the texture of the tissue accompanied with discernable changes in spatial arrangement of cells and their shapes, histology is an important method for differentiating tissue and cell phenotypes. In clinical fields, histopathology, which is a branch of histology, has been regularly used in diagnosis of diseases, especially in diagnosis of cancer [10], [11].

One of the important tasks in histology image research is to sort the images into different classes. In fact, the diagnosis itself is a process of binary classification. This task is usually quite challenging for the following reasons. First, the histology images are so called non-stationary because each region of the image can have different properties. Second, the variance of staining conditions and the difference among individuals increase the intra-class difference, which increases the difficulty for the classifier. Third, the inter-class difference is sometimes small, which is even hard for human experts to differentiate. In [12], the classification accuracies of three histology data sets, Liver Aging, Liver Gender, and Lymphoma are relatively low compared to most other data sets, which indicates the difficulty in classifying these images.

Many applications have been developed for classifying microscopic histology images. They generally fall into two categories. The first category is the application for general image classification. In fact, these applications target at a broad range of bio images so they are not designed to adapt to histology images. For example, Wndchrm [12] is an application of this type. In order to accommodate the characteristics of different data sets, this application includes a pool of 2659 different features for users to choose from and performs the further classification task. This application was

tested using the benchmark data sets from the IICBU Biological Image Repository [13]. Other open source applications for the research purpose in this category include CellProfiler [14] which focuses on cell image classification, and recently a support vector machine based Enhanced CellClassifier [15] is made available for the research community.

The second type of classification systems focuses on the specific task of histology image classification. Due to the importance of histopathology images in cancer diagnosis, benign and malignant cancer differentiation, and cancer cell classifications, the main research progress in this area lies in developing computer based automatic histology image classification systems targeting at the clinical fields. Within this research scope, there are generally two directions which most of the researchers follow. The first direction focuses on extracting specific features of certain data sets to represent the characteristics of the objects better. For example, the Adaptive Discriminant Wavelet Packet Transform was performed before extracting features to improve the robustness of the Meningioma Histology Images classification system [16]. In [17], Fractal Geometry based texture features were utilized to grade prostatic carcinoma. Another direction aims at performing a better segmentation of the image before further processing with the hope that the segmented objects (cell or nucleus) could represent the characteristics of the whole images better. Wang et al. indicated that the classification accuracies could reach 100% using a simple nearest neighbor classification algorithm on testing instances after segmenting the nuclei region from the original image and extracting features of chromatin patterns based on the segmented image [18]. A similar work in [19] also showed that extracting features from the gland and nuclei areas which were segmented from the original images gave promising results in prostate cancer grading, breast cancer detection, and breast cancer grading.

Though the aforementioned applications provide acceptable results for some data sets, many challenges still exist. First, the general applications are developed for microscopic image classification but not specifically for histology data sets, so they might not adapt to the characteristics of histology images. The relatively poor performance of Wndchrm on the aforementioned three histology data sets is an example [12]. On the other hand, feature extraction oriented development of classification applications is quite dependent on the specific dataset and sometimes requires domain knowledge, which is not desirable for a general histology classification system. At the same time, the segmentation based classification systems suffer from two problems which need to be solved. First, the segmentation process increases the computation cost. Second, it is still not uncommon to fail to segment an object within an image [7], which leads to information loss before feature extraction.

Collateral Representative Subspace Projection Modeling (C-RSPM) [20] is a supervised classification method developed in our previous work. It tries to train an array of Principal Component Classifiers (PCCs), each of which learns the distribution of training data of one class with an attempt to recognize the testing instances belonging to that class. Due to its relatively good performance and robustness, this classification method has been used in many domains including network intrusion detection [21] and multimedia concept detection [22].

In order to develop a general classification system for histology images, a C-RSPM based histology image classification framework is proposed and tested using three data sets from the IICBU-2008 benchmark suite of biological image data sets [13]. In this framework, the original image is first divided to equal-sized blocks. Further, the blocks having same location from different images are selected to form a block set. Low-level visual features are extracted from all the blocks and normalization is done within each block set. One C-RSPM model is trained on one block set and classifying the corresponding block from a testing image. An accuracy based voting scheme is developed to deduce the final label of each testing image. In this process, no object detection or segmentation method is needed and a wide range of features including color and texture features are used to help relax the dataset dependency. The classification results are compared with those of two other frameworks using seven classifiers, such as decision tree classifier (DT), support vector machine classifier (SVM), and so on. Overall, our proposed framework outperforms the two frameworks in the comparison and other reported results on the same data sets, which indicates the strength of the system.

This paper is organized as follows. In Section II, the proposed framework and detailed description of different components are presented. Section III shows the experiment results and our analysis. Section IV concludes the paper and discusses some future work.

II. THE PROPOSED FRAMEWORK

The overall framework is shown in Fig. 1. The proposed framework is composed of two components, namely *feature preparation* and *classification & fusion components.*

In the *feature preparation component*, each image is first divided to 25 equal-sized blocks. The reasons to divide the original image in this way will be introduced in Section II-A. The blocks in the same location from different images form a block set B_k, $1 \le k \le 25$. A set of 505 features, including color and texture features, are extracted from every block. The data are then split into a training set (two thirds of the whole data) and a testing set (one third of the whole data). It needs to be pointed out that the splitting is done for the original images but not for blocks in each block set. A specific normalization process described in Section II-B is used to normalize the data within each block set. After a Chi-Square based feature selection, the top fifty selected features are retained for further training and classification.

In the *classification & fusion component*, one C-RSPM model is built for each block set and therefore there are 25 models in total. The parameters in C-RSPM are selected automatically and adaptively based on the training dataset. Each block b_k $(1 \le k \le 25)$ from a testing image will be classified by the C-RSPM model built on the corresponding block set B_k. Since one testing image is composed of 25

blocks, there are 25 classification results correspondingly. The accuracy based Weighted Majority Voting Algorithm (WMVA) is proposed to perform late fusion and infer the final class label. The classification result is evaluated using accuracy, which is the percentage of the images in the testing dataset whose classification class labels determined by our proposed framework match the ground truth class labels. This is consistent with the results reported in the previous study [12] for the comparison purpose.

A. Image Division

There are three main reasons for dividing each image as proposed in our framework. First, it decreases space complexity for feature extraction [23]. Second, a previous study has demonstrated that when each image captures many cells, dividing the image into blocks to extract local features can, in some cases, improve the results in comparison with other approaches using global features [24]. Third, different regions of the original image may have different characteristics useful for classification. Therefore, the factor of locations is taken into consideration in this way.

The proposed scheme of image division is shown in Fig. 2. The original image (1388*1040) is divided to 16 Type I blocks (shown as the blocks marked by blue solid lines, from blocks 1 to 16) and 9 Type II blocks (marked by black dashed lines, from blocks 17 to 25). Both Type I and Type II blocks have the same size. As can be seen in Fig. 2, if all the 16 Type I blocks are joined together, the original image can be reconstructed. The reason why the type II blocks are included is to recover the information in the boundary area near the dividing lines between neighboring Type I blocks. The information in these areas might be missing if only Type I blocks are used. Of course, the total number of blocks to divide an image can be modified as needed.

Since a general microscopic classification system for histology image is proposed, a wide range of visual features are extracted. As in our previous work [25], a set of 505 visual features are extracted from each block (8 features are excluded because they are for face detection). The visual features include: Color Dominant (16 features), Color Histogram (51 features), Color Moment (108 features), Edge Histogram (47 features), Texture Co-occur (36 features), Texture Wavelet (219 features), Texture Tamura (3 features), Texture Gabor (24 features), and Local Binary Patterns (1 feature). In order to reduce the scaling factor for numeric values, the Z score normalization method is applied to the data using (1).

$$Z = \frac{W_k - \mu_k}{\sigma_k}, \qquad (1)$$

where W_k is the raw training data, μ_k is the mean of the training data, σ_k is the standard deviation of the training data. Here, the normalization is done within each block set B_k, $1 \le k \le 25$ so that the location of the blocks in each B_k is taken into consideration. In addition, this normalization method also helps the feature selection step by keeping the discriminatory capability of the features.

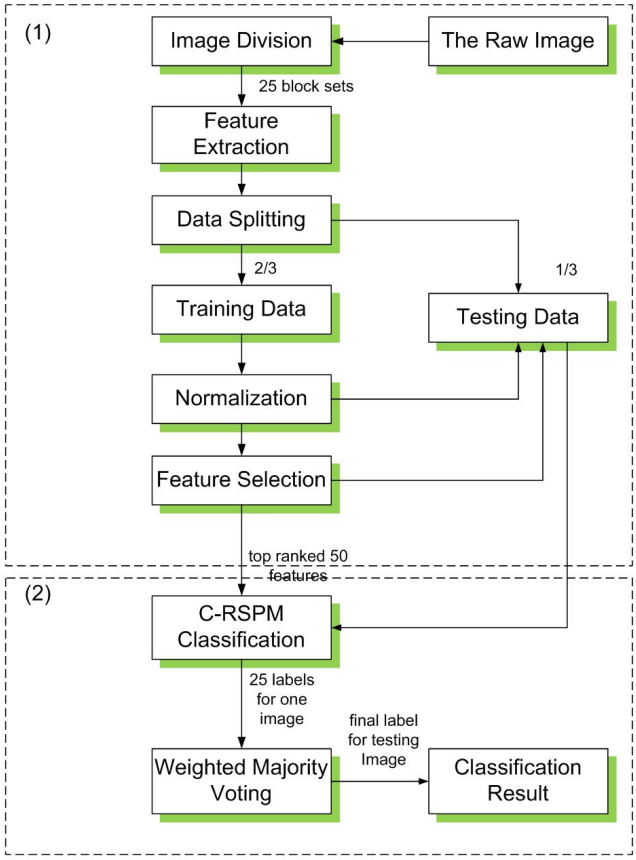

Figure 1. System architecture of the proposed classification framework

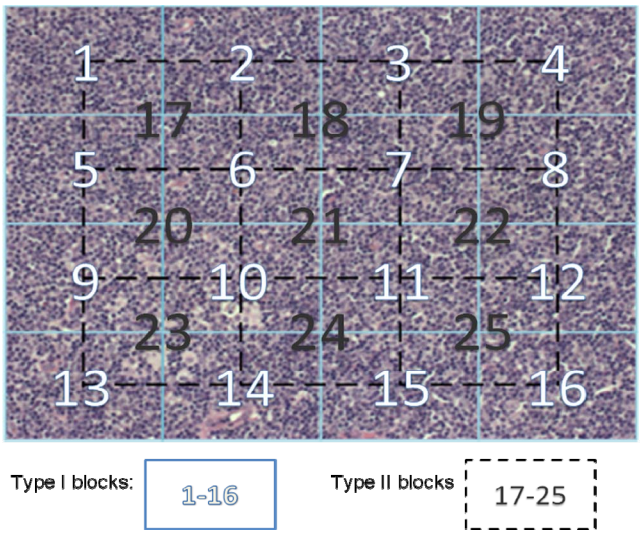

Figure 2. Division of the image to 25 blocks

Since feature selection is not the focus of this paper, the Chi-Square (χ^2) feature selection algorithm available in Weka [26] is applied. The Chi-Square measure evaluates features by ranking the Chi-Square statistics of each feature with respect to the class label. After ranking of the selected features, only the top 50 features are used in building the classification model, which only requires one tenth of the computation load when using the whole feature set.

B. C-RSPM Classifier

The C-RSPM Classifier is used to build the K (where K=25 in this study) classification models. The general idea of the C-RSPM Classifier is to build multiple Principal Component Classifiers (PCCs), where the number of PCCs is determined by the number of the classes in an application. Each PCC is trained to learn the similarities among the training data instances of a particular class and used to recognize the testing data instances that are normal to that class. For the training data of a certain class, let i = 1, 2,…, N, j = 1, 2,…, T, and $\mathbf{X} = \{x_{ij}\}$ be a (N x T)-dimensional matrix with N training data instances and T features. Let $X_i = [x_{i1}\, x_{i2} \ldots x_{iT}]'$ be the column vector representing the features for the i^{th} data instance. In order to identify the most representative instances in the training set for a certain class, outliers are removed by calculating the Mahalanobis distance using (2).

$$d_i^2 = (X_i - \overline{X})' S^{-1} (X_i - \overline{X}), \quad (2)$$

where $\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$ and

$$S = \frac{1}{N-1}\sum_{i=1}^{N} (X_i - \overline{X})(X_i - \overline{X})'.$$

The data trimming rate (γ) parameter is set to trim the instances corresponding to the ($\gamma \cdot N$) largest distance values. After data trimming, the new matrix having L instances (L<N) and T features with elements x_{ij}, where i = 1, 2, …, L and j = 1, 2, ..., T is normalized to get $\mathbf{Z}$ and the correlation matrix $\mathbf{M}$ of $\mathbf{Z}$ is calculated. Afterwards, T eigenvalue-eigenvector pairs (λ_1, E_1), (λ_2, E_2),…, (λ_T, E_T) are computed for $\mathbf{M}$. Then the matrix $\mathbf{Z}$ is projected to the T-dimensional eigenspace composed of all T principal components from $\mathbf{M}$ to get $\mathbf{Y} = \{y_{ij}\}$ in which the row vector $Y_i = [y_{i1}\, y_{i2}, \ldots, y_{iT}]$ is the projection of each normalized training data instance in $\mathbf{Z}$ onto the eigenspace. Then the principal components whose corresponding eigenvalues are greater than the threshold $\boldsymbol{\eta}$ are kept as the representative principal components. Let $\mathbf{F} = \{f_q\}$, f_q represents the indexes of the features corresponding to the representative principal components in $\mathbf{Y}$. A distance vector $C = \{c_i\}$, i = 1, 2, …, L is computed using (3).

$$c_i = \sum_{f \in F} \frac{(y_{if})^2}{\lambda_f}. \quad (3)$$

Afterwards, the elements of C are sorted in the ascending order and only the $\boldsymbol{\beta}$ smallest values are retained. Assuming $\boldsymbol{\alpha}$ is the false alarm rate, $\boldsymbol{\beta}$ is calculated as the value of the nearest integer to which the value of ($(1\text{-}\boldsymbol{\alpha}) \cdot L$) is rounded. The threshold value c_{thres} is set to the maximum value of the retained elements in C. For each testing data instance (i.e., a testing image), let $X' = [x'_1\, x'_2 \ldots x'_T]'$ be the column vector representing the features for that testing instance. After normalization, the projection vector $Y' = [y'_1\, y'_2 \ldots y'_T]$ is commuted by projecting X' onto the same T-dimensional eigenspace composed of all T principal components from $\mathbf{M}$. c' value is calculated using (4).

$$c' = \sum_{f \in F} \frac{(y'_{if})^2}{\lambda_f}. \quad (4)$$

This testing data instance is classified as statistically negative to the class of the training data if $c' > c_{thres}$ and classified as positive to the class if $c' \leqslant c_{thres}$. Theoretically, a testing data instance is recognized by one and only one classifier as positive instance. However, in practice, the testing data instance could be rejected by all PCCs or it might be accepted by more than one PCC. Therefore, an ambiguity solver module is included to solve this problem. The specific details of the classifier could be found in [22]. The parameters such as γ, $\boldsymbol{\eta}$ and $\boldsymbol{\alpha}$ used in the proposed framework are tuned adaptively and automatically based on the training data.

C. The Weighted Majority Voting Alogrithm (WMVA)

After building the K (where K=25 in this study) models for the 25 block sets, the testing images are classified using these models. Specifically, each testing image is first divided to 25 blocks using the same scheme described in Section II-A, and each block is classified using the corresponding C-RSPM model for that block set. Therefore, the whole image gets 25 labels and a final decision needs to be made. Rather than considering each label equally, the weight of the voting score of each block is computed based on the accuracy of the model.

Let G_1, G_2... G_K be the K models and D_1, D_2...D_K be the classification results of the blocks of one testing image. Suppose that there are U classes in total in a certain data set, labeled as Class 1, Class 2 , … , Class U correspondingly. The matrix $\mathbf{R} = \{r_{ho}\}$, $1 \leq h \leq K$, $1 \leq o \leq U$ is the weight matrix, where r_{ho} is the weight of Model h for Class o. The final score for Class o is calculated using (5), where H is the set of those elements with each element w being the index of the model whose classification result D_w=Class o. The testing image is classified to the label which gets the highest score.

$$Score(Class\ \ o) = \sum_{h \in H} r_{ho}, \quad (5)$$

where $H = \{w \mid D_w = Class\ \ o, 1 \leq w \leq K\}$

The simplest scoring scheme is to assign 1 to all the elements in matrix $\mathbf{R}$, which is the majority voting scheme.

In this framework, **R** is calculated based on the classification accuracy. Let the total number of blocks be K and the total number of classes be U. In our study, $K = 25$ and U depends on the data sets used. The pseudo code for computing the weight is presented as follows. In this way, all the scores are normalized to the range of 0 to 1. In addition, the differences among the accuracy are amplified so the accuracy of the models could play a significant role in the majority voting scheme. Experimental results show that using the WMVA gives better performance than using the simple majority voting algorithm.

WEIGHT-GENERATION (K, U)

```
1    for model_h ← model_1 to model_K
2      for class_o ← class_1 to class_U
3          NumOfTP(h,o) = Number of true positive
           instances in the training data set for class_o
           of model_h
4          NumClass(o) = Number of ground truth
           instances for class_o in the training data set
5          AccuracyM(h,o) = NumOfTP(h,o)/ NumClass(o)
6      next
7    next
8    SMax = Maximum(AccuracyM(h,o))
9    SMin = Minimum(AccuracyM(h,o))
10   SDif = SMax - SMin
11   for model_h ← model_1 to model_K
12     for class_o ←class_1 to class_U
13         R (h,o) = (AccuracyM(h,o)-SMin)/SDif
14     next
15   next
```

III. Experiments and Results

The proposed framework is validated using three benchmark histology data sets from the IICBU Biological Image Repository [13]. The first two data sets are for cell biology research and the third one is for clinical diagnosis. These three data sets are described as follows. In our study, for each data set, 2/3 of the data instances in each class are randomly selected as the training data set to train the classifier and the remaining 1/3 of the data instances are used as the testing data set to validate the classifier.

Liver Aging: The data set contains 528 microscopy images (1388x1040) of tissues of four classes. They are tissues from 1 month female mice on ad-libitum diet (99 images), 6 month female mice on ad-libitum diet (115 images), 16 month female mice on ad-libitum diet (162 images), and 24 month female mice on ad-libitum diet (152 images). Livers were extracted from the sacrificed mice and the tissue was sectioned and stained using H&E staining method. The sample images are shown in Fig. 3. It should be pointed out that although the images are prepared by one researcher, which could help control the staining variance, the training set of images are selected randomly so that they may not come from the same animal, which could increase the inner-class difference [13].

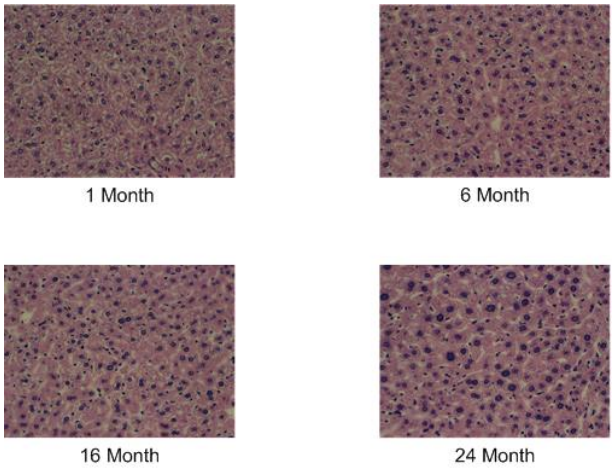

Figure 3. Sample images in Liver Aging data set

Liver Gender: The data set contains 265 images (1388x1040) of two classes. They are liver tissue images from 6 month male mice on ad-libitum diet (150 images) and 6 month female mice (115 images) on ad-libitum diet. They are prepared in the same way as the Liver Aging image data set. The sample images are shown in Fig. 4.

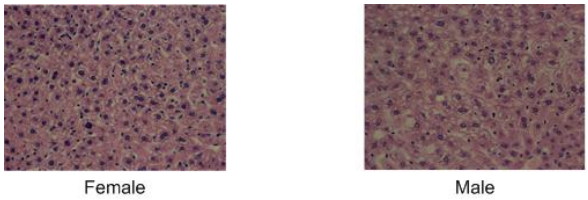

Figure 4. Sample images in Liver Gender data set

Lymphoma: Malignant lymphoma is a cancer affecting lymph node. In this data set, 376 images (1388x1040) of three types of malignant lymphoma are represented. They are chronic lymphocytic leukemia (CLL) cells (114 images), follicular lymphoma (FL) cells (139 images), and mantle cell lymphoma (MCL) cells (123 images). The sample images are shown Fig. 5. These images are from biopsies of different patients in different hospitals, so the inner class difference could be even larger than the previous two data sets because of the staining variance [13].

As shown in the sample images, the task is quite challenging since there are no obvious visual markers even for human eyes. To show the effectiveness of the proposed framework, the performance is compared with that of two other fusion methods using common classifiers in Weka [26], including C4.5, JRIP, AdaBoost (with C4.5), k-Nearest Neighbors (k=3), Support Vector Machine (with poly-kernel), Bayes-Net, and Naïve-Bayes applications. The two fusion frameworks are called early fusion and late decision. In early fusion framework, the feature set for each image is

summarized in the matrix V = $\{v_{qp}\}$, $1 \leqslant q \leqslant 25$, $1 \leqslant p \leqslant 50$, where $V_q = [v_{q1}\ v_{q2} \cdots v_{q50}]$ is the row vector of feature values for block q. A new feature vector $V' = [v'_p]$, $1 \leq p \leq 50$ is computed in (8). The new feature vector V' is used for classification directly.

$$v'_p = (\sum_{q=1}^{25} v_{qp})/25. \quad (6)$$

The late decision framework is to perform the classification using the blocks directly. After getting the class label of each block, a majority voting scheme is used to deduce the final label of the image. The evaluation criterion used is the classification accuracy.

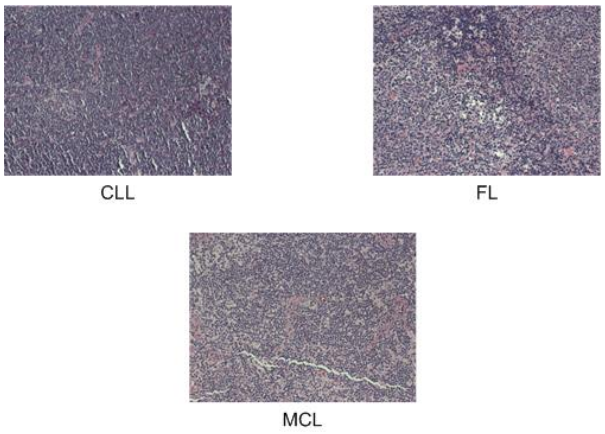

Figure 5. Sample images in Lymphoma data set

TABLE I.

CLASSIFICATION ACCURACY FOR LIVER AGING DATA SET

TABLE I (A)	Published	C-RSPM+ WMVA
Accuracy	51.00%	**96.41%**

TABLE I (B)	Early Fusion	Late Decision
C4.5	88.08%	87.50%
JRIP	84.87%	86.77%
AdaBoost	94.69%	87.50%
3NN	95.09%	87.70%
SVM	94.70%	85.23%
BayesNet	79.41%	73.67%
NaiveBayes	66.46%	62.87%
Average	86.19%	81.61%

Three cross validation was conducted to evaluate the framework and the average accuracies for the three data sets are shown in Table I, Table II, and Table III, correspondingly. In each table, table (A) shows the published highest accuracy in the column "Published" and the accuracy of the proposed framework in the column "C-RSPM+WMVA". In table (B), the columns "Early Fusion" and "Late Decision" show the classification accuracies of several common classifiers in the early fusion framework and late decision framework, correspondingly.

Generally speaking, the classification accuracies in the liver aging dataset and the liver gender dataset are higher than the Lymphoma dataset, which matches with the fact that the larger staining variance in the third dataset could increase the intra-class difference thus confusing the classifiers. The performance of the classifiers for the second dataset is better than the first one because it is a two-way classification problem which is usually simpler for the classifier than the three-way classification problem for the first data set.

TABLE II

CLASSIFICATION ACCURACY FOR LIVER GENDER DATA SET

TABLE II(A)	Published	C-RSPM+ WMVA
Accuracy	69.00%	**99.24%**

TABLE II(B)	Early Fusion	Late Decision
C4.5	85.30%	90.95%
JRIP	82.26%	87.17%
AdaBoost	88.32%	92.08%
3NN	90.58%	90.95%
SVM	88.33%	89.06%
BayesNet	86.45%	81.90%
NaiveBayes	76.99%	75.87%
Average	85.46%	86.85%

It can be also seen that the proposed framework constantly outperforms other frameworks using common classifiers. For example, for the three data sets, k-nearest neighbor classifier (k=3) using early fusion performs the second best for the first data set and Adaboost (with C4.5) classifier performs the second best for the second data set in early fusion framework. The k-nearest neighbor classifier sometimes identifies the same animal by finding the most similar instances in the training set for a testing instance. Therefore, it performs relatively well. The Adaboost (with C4.5) performs well for the second data set because this algorithm is better at binary classification problems [27]. It could be seen that the proposed framework improves the final classification results. It builds a model on each block set so that the information of the locations of the blocks in the image is taken into consideration, which is very important because different blocks of the histology images are different internally in terms of histology structure and distribution of cells. On the other hand, the late decision and early fusion frameworks, in which all the blocks from one

image are treated the same, lose the information of the locations. It should also be pointed out that the reported highest average classification accuracies using the same data sets are 51%, 69%, and 85% in [12] for those three data sets. The average classification accuracies of our proposed framework using the weighted majority voting algorithm are 96.40%, 99.24%, and 92.70% correspondingly, which validates the effectiveness of the proposed framework. In addition, the proposed framework neither depends on extracting specific features for certain image sets nor requires a segmentation method, and the testing results on the histology images from both research and clinical domains prove its prospective broad applications

TABLE III

CLASSIFICATION ACCURACY FOR LYMPHOMA DATA SET

TABLE III(A)	**Published**	**C-RSPM+ WMVA**
Accuracy	85.00%	**92.70%**

TABLE III(B)	**Early Fusion**	**Late Decision**
C4.5	75.78%	72.77%
JRIP	75.78%	73.83%
AdaBoost	83.05%	76.50%
3NN	84.79%	76.73%
SVM	67.14%	75.97%
BayesNet	66.03%	67.73%
NaiveBayes	61.31%	59.97%
Average	73.41%	71.93%

IV CONCLUSION AND FUTURE WORK

This paper presents a novel histology image classification framework with the C-RSPM classifier and a weighted majority voting algorithm for late fusion. The images are first divided into 25 blocks according to different locations. After feature extraction, normalization, and feature selection, 25 C-RSPM models are built for 25 block sets. Finally, the classification results of the testing blocks from one image are used in the weighted majority voting algorithm to infer the final class label of the image. The experiments with three benchmark histology data sets from the IICBU Biological Image Repository show that the proposed framework constantly outperforms other frameworks and achieves better results than the previously published ones. Moreover, the proposed framework is not data set dependent or image segmentation dependent, and therefore it reduces the computation complexity and could be used for other general histology image classification easily.

For the future work, more histology data sets need to be tested using the proposed framework and heuristic block division algorithm, such as multi-resolution division algorithm, will be investigated to improve the classification results. The final goal is to build a broad and general histology image classification and retrieval system for not only helping the biologists to better analyze experimental results but also assisting the pathologists in the clinical field for disease diagnosis.

ACKNOWLEDGMENT

For Shu-Ching Chen, this work was supported in part by NSF HRD-0833093.

REFERENCES

[1] Roy Wollman and Nico Stuurman, “High throughput microscopy: from raw images to discoveries,” Journal of Cell Science, Vol. 120, No. 21, pp. 3715–3722, November 2007.

[2] Xi Long, W. Louis Cleveland, and Y.Lawrence Yao, “Multiclass detection of cells in multicontrast composite images,” Computers in Biology and Medicine, Vol. 40, No. 2, pp. 168-178, February 2010.

[3] Yue Huang Xuezhi Sun, and Guangshu Hu, “An automatic integrated approach for stained neuron detection in studying neuron migration,” Microscopy Research and Technique, Vol. 73, No. 2, pp. 109-118, February 2010.

[4] Min Hyeok Bae, Rong Pan, Teresa Wu, and Alexandra Badea, “Automated segmentation of mouse brain images using extended MRF,” Neuroimage, Vol. 46, No. 3, pp. 717-725, July 2009.

[5] H.T. Madhloom, S.A. Kareem, H. Ariffin, A.A. Zaidan, H.O. Alanazi, and B.B. Zaidan, “An automated white blood cell nucleus localization and segmentation using image arithmetic and automatic threshold,” Journal of Applied Sciences, Vol. 10, No. 11, pp. 959-966, 2010

[6] Reiko Minamikawa-Tachino, Noriko Kabuyama Toshiyuki Gotoh, Seiichiro Kagei, Masatoshi Naruse, Yasutomo Kisu Takushi Togashi, Sumio Sugano, Hitohide Usami, and Nobuo Nomura“High-throughput classification of images of cells transfected with cDNA clones,” Comptes Rendus Biologies, Vol. 326, No.10-11, pp. 993-1001,October-November 2003.

[7] Loris Nanni and Alessandra Lumini, “A reliable method for cell phenotype image classification,” Artificial Intelligence in Medicine, Vol. 43, No. 2, pp. 87-97, June 2008.

[8] Hanchuan Peng, “Bioimage informatics: a new area of engineering biology,” Bioinformaics, Vol 24, No.17, pp. 1827-1836, September 2008.

[9] Ralph D. Lillie and Harold M. Fullmer, Histopathologic Technic and Practical Histochemistry , 4th ed .New York:McGraw-Hill, 1976

[10] Stephen Berns and Gary Pearl, “Review of pineal anlage tumor with divergent histology,” Archives of Pathology & Laboratory Medicine, Vol. 130, No. 8, pp.1233-1235,August 2006.

[11] Pranab Dey, “Cancer nucleus: morphology and beyond,” Diagnostic Cytopathology, Vol. 38, No. 5,pp.382-390,May 2010.

[12] Lior Shamir, Nikita Orlov, David Mark Eckley, Tomasz Macura, Josiah Johnston, and Ilya G Goldberg, “Wndchrm-an open source utility for biological image analysis,” Source Code for Biology and Medicine, Vol. 3, No. 13, July 2008.

[13] Lior Shamir, Nikita Orlov, David Mark Eckley, Tomasz J. Macura, and Macura,Ilya G. Goldberg, “IICBU 2008 - A proposed benchmark suite for biological image analysis,” Medical and Biological Engineering and Computing, Vol. 46, No. 9, pp. 943-947,September, 2008.

[14] Anne E Carpenter, Thouis R Jones, Michael R Lamprecht, Colin Clarke, In H Kang, Ola Friman, David A Guertin, Joo H Chang, Robert A Lindquist, Jason Moffat, Polina Golland, and David M

Sabatini, “CellProfier: image analysis software for identifying and quantifying cell phenotypes,” Genome Biology, Vol. 7, No. 10, pp.R100, 2006

[15] Benjamin Misselwitz, Gerhard Strittmatter, Balamurugan Periaswamy, Markus C. Schlumberger, Samuel Rout, Peter Horvath, Karol Kozak, and Wolf-Dietrich Hardt, “Enhanced cellClassifier: a multi-class classification tool for microscopy images,”BMC Bioinformatics, Vol. 11, No. 30, 2010

[16] Hammad Qureshi, Nasir Rajpoot, Tim Nattkemper, and Volkmar Hans,“A robust adaptive wavelet-based method for classification of Meningioma histology images,” Proceedings MICCAI'2009 Workshop on Optical Tissue Image Analysis in Microscopy, Histology, and Endoscopy (OPTIMHisE), London (UK), 2009, in press.

[17] Po-Whei Huang and Cheng-Hsiung Lee, “Automatic classification for pathological prostate images based on fractal analysis,”IEEE Transactions on Medical Imaging, Vol. 28, No. 7, pp.1037-1050, July 2009.

[18] Wei Wang, John A.Ozolek, and Gustavo K. Rohde, “Detection and classification of thyroid follicular lesions based on nuclear structure from histopathology images,” Cytometry Part A, Vol. 77A, No. 5, pp. 485-494, May 2010.

[19] S. Naik, S. Doyle, S. Agner, A. Madabhushi, M. Feldman, and J. Tomaszewski, “Authomated gland and nuclei segmentation for grading of prostate and breast cancer histopathology”, 5th IEEE International Symposium on Biomedical Imaging, pp.1037-1050,May 14-17, Paris, France, 2008.

[20] Thiago Quirino, Zongxing Xie, Mei-Ling Shyu, Shu-Ching Chen, and LiWu Chang, “Collateral representative subspace projection modeling for supervised classification,” Proceedings of the 18th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'06), pp. 98-105, November 13-15, Washington D.C., USA.

[21] Varsha Sainani and Mei-Ling Shyu, “A hybrid layered multiagent architecture with low cost and low response time communication protocol for network intrusion detection systems,” The IEEE 23rd International Conference on Advanced Information Networking and Applications (AINA-09), pp. 154-161, Bradford,UK, May 26-29, 2009.

[22] Shu-Ching Chen, Mei-Ling Shyu, and Min Chen, “An effective multi-concept classifier for video streams,” Proceedings of the Second IEEE International Conference on Semantic Computing(IEEE ICSC2008), pp. 80-87, Santa Clara, CA, USA, August 4-7, 2008.

[23] Jason D. Browning and Steven L. Tanimoto, “Segmentation of pictures into regions with a tile-by-tile method,” Pattern Recognition, Vol. 15, No. 1, pp. 1-10, 1982

[24] Lior Shamir, D Mark Eckley, and Ilya G Goldberg, “Image tiling vs. cell segmentation - a case study,” 47th American Society for Cell Biology Meeting (ASCB'07). pp. 35. Washington, D.C., December 2007.

[25] Lin Lin, Chao Chen, Mei-Ling Shyu, Fausto Fleites, and Shu- Ching Chen, "Florida International University and University of Miami TRECVID 2009 - high level feature extraction," TRECVID workshop 2009.

[26] Ian H. Witten and Eibe Frank, Data Mining Practical Maching Learning Tools and Techniques, 2nd ed. San Francisco: Morgan Kaufmann,2005.

[27] Ji Zhu, Hui Zou, Saharon Rosset and Trevor Hastie, “Multi-class AdaBoost,” Statistics and Its Interface, Vol. 2, No. 3, pp. 349-360, 2009