Spring 2018

Data fusion techniques for biomedical informatics and clinical decision support

Peng Guo

Department: Electrical and Computer Engineering

DATA FUSION TECHNIQUES FOR BIOMEDICAL INFORMATICS AND CLINICAL DECISION SUPPORT

by

PENG GUO

A DISSERTATION

Presented to the Graduate Faculty of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

In Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

in

ELECTRICAL ENGINEERING

2018

Approved by

R.J. Stanley, Advisor
R.H. Moss
W.V. Stoecker
D.C. Wunsch
B. Shrestha
V.A Samaranayake

PUBLICATION DISSERTATION OPTION

This dissertation consists of the following three articles which have been submitted for publication as follows:

Paper I: Pages 6-46 have been published JBHI Journal.

Paper II: Pages 47-76 have been published by JPI Journal.

Paper III: Pages 77-103 have been published by KEI Journal.

ABSTRACT

Data fusion can be used to combine multiple data sources or modalities to facilitate enhanced visualization, analysis, detection, estimation, or classification. Data fusion can be applied at the raw-data, feature-based, and decision-based levels. Data fusion applications of different sorts have been built up in areas such as statistics, computer vision and other machine learning aspects. It has been employed in a variety of realistic scenarios such as medical diagnosis, clinical decision support, and structural health monitoring. This dissertation includes investigation and development of methods to perform data fusion for cervical cancer intraepithelial neoplasia (CIN) and a clinical decision support system. The general framework for these applications includes image processing followed by feature development and classification of the detected region of interest (ROI). Image processing methods such as k-means clustering based on color information, dilation, erosion and centroid locating methods were used for ROI detection. The features extracted include texture, color, nuclei-based and triangle features. Analysis and classification was performed using feature- and decision-level data fusion techniques such as support vector machine, statistical methods such as logistic regression, linear discriminant analysis and voting algorithms.

ACKNOWLEDGMENTS

I would like to sincerely thank my advisor Prof. Joe Stanley for providing me with an opportunity of participating in graduate studies and research under his guidance. I should thank him for his constant support and the flexibility. He allowed me to pursue my interests in my study and research for the last three years. He has always been encouraging me to seek the answer with considerable ideas and passion. I also wish to thank my committee members Dr. Moss, Dr. Wunsch, Dr. Shrestha and Dr. Stoecker for taking time out of their schedule and agreeing to serve in my committee. For the great internship experience I had at National Institutes of Health, many thanks are due to Rodney Long , Sameer Antani and Jaylene Xue. They are the best researchers and scientists I have ever met and they always helped me stay focused and overcome any difficulties that I have ever come across. Finally, and most importantly, I am extremely grateful to my family, who have always wished the best for me and showered their great love and care on me. I would like to thank my mom for her most considerable support and love which gives me courage and faith to move on. And I also need to thank my cousin Jiahang Han and my aunt Changxin for their time with me in the past.

TABLE OF CONTENTS

Page

PUBLICATION DISSERTATION OPTION iii

ABSTRACT iv

ACKNOWLEDGMENTS v

LIST OF ILLUSTRATIONS x

LIST OF TABLES xiii

SECTION

1. INTRODUCTION 1
   - 1.1. OVERVIEW OF DATA FUSION 1
   - 1.2. PROBLEM DESCRIPTION 2
   - 1.3. SUMMARY OF CONTRIBUTIONS 3
     - 1.3.1. Improved Localized Classification of Cervical Intraepithelial Neoplasia (CIN) Diagnosis 3
     - 1.3.2. Fusion Paradigm Development Based on Inter/Intra-pathologist CIN Labeling for Localized and Entire Image CIN Classification in Digitized Histology Images 4
     - 1.3.3. Advanced Features Development to Find Image for Application to Clinical Decision Support 4

PAPER

I. NUCLEI-BASED FEATURES FOR UTERINE CERVICAL CANCER HISTOLOGY IMAGE ANALYSIS WITH FUSION-BASED CLASSIFICATION 6

   ABSTRACT 6

1. INTRODUCTION 7
2. METHODS 9
2.1. Medial Axis Detection and Segments Creation 10
2.2. Feature Extraction 14
2.2.1. Texture and Cellular Features 14
2.2.2. Nuclei Features 16
2.2.3. Acellular Features 22
2.2.4. Triangle Features 25
3. EXPERIMENT PERFORMED 27
3.1. Fusion-Based CIN Grade Classification of Vertical Segment Images 27
3.1.1. Classification of the Whole Epithelium 29
3.1.2. Feature Evaluation and Selection 30
4. EXPERIMENTAL RESULTS AND ANALYSIS 31
4.1. Experimental Results 31
4.2. Analysis of Results 36
5. SUMMARY AND FUTURE WORK 41
6. ACKNOWLEDGMENTS 42
REFERENCES 43

II. ENHANCEMENTS IN LOCALIZED CLASSIFICATION FOR UTERINE CERVICAL CANCER DIGITAL HISTOLOGY IMAGE ASSESSMENT 47
ABSTRACT 47
1. INTRODUCTION 48
2. METHODS 50
2.1. Pathologist Ground Truth Description 51
2.2. Medial Axis Detection and Segments Creation 54
2.3. Feature Extraction 54

2.3.1. Texture and Color Features 54
2.3.2. Nuclear Features 54
2.3.3. Acellular Features 56
2.3.4. Combination Features 57
2.3.5. Triangle Features 57
3. EXPERIMENT PERFORMED 58
3.1. Fusion-based Cervical Intraepithelial Neoplasia Grade Classification of Vertical Segment Images Labeling 58
3.2. Feature Evaluation and Selection 60
4. EXPERIMENTAL RESULTS AND ANALYSIS 61
4.1. Experimental Results 61
4.2. Analysis of Results 64
5. CONCLUSION 72
6. ACKNOWLEDGMENTS 73
7. FINANCIAL SUPPORT AND SPONSORSHIP 73
8. CONFLICTS OF INTEREST 73
REFERENCES 74

III. FEATURES ADVANCES TO AUTOMATICALLY FIND IMAGES FOR APPLICATION TO CLINICAL DECISION SUPPORT 77
ABSTRACT 77
1. INTRODUCTION 78
2. METHODOLOGY 82
2.1. Data Set Examined 82
2.2. Features and Feature Groups Investigated 83
2.3. Features and Feature Groups Investigated 83
2.3.1. HSV Histogram Correlation Features 83
2.3.2. Group 3 87

3. EXPERIMENTS PERFORMED 88
3.1. Benchmark Features 88
3.2. Nearest Neighbor Classifier 89
4. EXPERIMENTAL RESULTS AND DISCUSSION 90
5. CONCLUSION 94
6. ACKNOWLEDGEMENT 94
REFERENCES 95

SECTION

2. SUMMARY AND CONCLUSIONS 104

REFERENCES 106

VITA 121

LIST OF ILLUSTRATIONS

Figure Page

PAPER I

1. CIN grade label examples highlighting the increase of immature atypical cells from epithelium bottom to top with increasing CIN severity. (a) Normal. (b) CIN 1. (c) CIN 2. (d) CIN 3. 7

2. Example of incorrect medial axis determined using distance transform only (solid line). The desired medial axis is manually drawn and is overlaid on the image (dashed line). The left-hand, right-hand, and interior sections are labeled on the bounding box image to highlight the limitations of the distance transform algorithm. 10

3. Bounding box of epithelium with control points labeled. 11

4. Bounding box partitioning with masks combinations shown based on control points from Fig. 3 as a part of epithelium orientation determination algorithm. (a) Mask 1 and Mask 2. (b) Mask 3 and Mask 4. (c) Mask 7 and Mask 8. (d) Mask 5 and Mask 6. 12

5. Epithelium image example with vertical segment images (I_1 ,I_2, I_3 ,..., I_{10}) determined from bounding boxes after dividing the medial axis into ten line segment approximations after medial axis computation. 13

6. Sample shading representatives within epithelium image used for determining cellular features. 16

7. Example of image preprocessing. (a) Original luminance image I. (b) Sharpened image $I_{sharpen}$ obtained after average filtering of I. 17

8. Example of image processing steps to obtain nuclei cluster pixels from K-means algorithm from histogram equalized image. (a) Histogram equalized image determined from Figure. 7(b). (b) Mask image obtained from K-means algorithm with pixels closest to nuclei cluster. 18

9. Image examples of nuclei detection algorithm. (a) Image with preliminary nuclei objects obtained from clustering [step 1-Figure 8(c)]. (b) Image closing to connect nuclei objects (step 2). (c) Image with hole filling to produce nuclei objects (step 3). (d) Image opening to separate nuclei objects (step 4). (e) Image with nonnuclei (small) objects eliminated (step 5). 19

10. Example L* image for light area detection. 21

11. Adaptive histogram equalized image of Fig. 10. 21

12. Thresholded image of Fig. 11. 21

13. Example image of light area clusters after K-means clustering. 21

14. Example morphological dilation and final light area mask. (a) Morphological dilation and erosion process after K-means clustering. (b) Final light area mask, after eliminating regions with areas smaller than 100 pixels..... 22

15. Example of nuclei detection comparison between the circular-Hough method and the method presented in this paper. (a) Original vertical segment. (b) Example of circular-Hough method; note the nuclei misses and false detections. (c) Nuclei detected using the algorithm from Section II..... 25

16. Distribution of nuclei centroids as vertices for DTs in bottom layer (green), mid layer (red), and top layer (blue). 26

17. DTs in bottom layer (green lines), mid layer (red lines), and top layer (blue lines). 27

18. Fusion-based approach versus whole image approach..... 29

19. Example image of nuclei detection over whole image without creating vertical segments; the top image is the original epithelium image; bottom is the nuclei mask of this image..... 30

20. Misclassification example of a CIN2 image labeled as a CIN3. 37

21. Misclassification example of a CIN3 image labeled as a CIN1. 37

PAPER II

1. Cervical intraepithelial neoplasia grade label examples highlighting the increase of immature atypical cells from epithelium bottom to top with increasing cervical intraepithelial neoplasia severity. (a) Normal, (b) cervical intraepithelial neoplasia 1, (c) cervical intraepithelial neoplasia 2, (d) cervical intraepithelial neoplasia 3 49

2. Digitized pathology epithelium image analysis procedures. 51

3. Epithelium image example with vertical segment images (I_1, I_2, I_2,...,I_{10}) determined from bounding boxes after dividing the medial axis into ten line segment approximations after medial axis computation..... 52

4. Image examples of nuclei detection algorithm. (a) Image with preliminary nuclei objects obtained from clustering (Step 1). (b) Image closing to connect nuclei objects (Step 2). (c) Image with hole filling to produce nuclei objects (Step 3). (d) Image opening to separate nuclei objects (Step 4). (e) Image with nonnuclei (small) objects eliminated (Step 5) 55

5. Misclassification example of a cervical intraepithelial neoplasia 2 image labeled as a cervical intraepithelial neoplasia 3..... 56

6. Misclassification example of a cervical intraepithelial neoplasia 2 image labeled as a cervical intraepithelial neoplasia 1..... 58

PAPER III

1. Image examples (positive id) from each of the category numbers listed in Table 1. (a) Doppler ultrasound image (reproduced with permission [52]). (b) CT image with emphysema (reproduced with permission [53]). (c) Knee x- ray image (reproduced with permission [54]). (d) Mediastinal CT image (reproduced with permission [55]). (e) Abdominal CT image showing liver blood vessels (reproduced with permission [56]). (f) Chest CT image showing micro nodules (reproduced with permission [57]). (g) X-ray image containing one or more fractures (reproduced with permission [58]). (h) CT liver abscess (reproduced with permission [59]). (i) MRI or CT of colonoscopy (reproduced with permission [60]). (j) Photograph of tumor (reproduced with permission [61]). (k) Image of muscle cells (reproduced with permission [62]). (l) Image containing a Budd-Chiari malformation (reproduced with permission [63]). (m) Gastrointestinal neoplasm (reproduced with permission [64]). (n) Pulmonary embolism all modalities (reproduced with permission [65]). 84

2. The WDD functions used for computing correlation-based features with the HSV unsmoothed and smoothed histograms and the Gabor filter coefficients (reproduced with permission [1]) 86

LIST OF TABLES

Table Page

PAPER I

1. Feature description 15
2. Confusion matrix results for fusion-based classification using all 27 features (F1-F27) for SVM and LDA classifiers for both experts 31
3. Confusion matrix results for whole image classification using all 27 features (F1-F27) for SVM and LDA classifiers for both experts 32
4. Confusion matrix results for fusion-based classification using reduced features for SVM and LDA classifiers for both experts 33
5. Features with corresponding p-values and AIGR 34
6. CIN discrimination rates for fusion-based classification using all features, whole image classification, and reduced feature set fusion-based classification for both experts 35
7. Summary of classification accuracies: previous research versus reduced feature set results in this paper 38

PAPER II

1. Ground truth cervical intraepithelial neoplasia grade labels for both experts 53
2. Individual vertical segment exact class label classification results based on all 27 features using same expert labels for training-testing sets (RZ-RZ and SF-SF) 61
3. Features with corresponding p-values and attribute information gain ratio 62
4. Fusion-based whole image percentage correct cervical intraepithelial neoplasia discrimination rates using all features using the same expert for training and testing sets 64
5. Fusion-based whole Image percentage correct cervical intraepithelial neoplasia discrimination rates using reduced features with the same expert for training and testing sets 64
6. Best confusion matrix results for fusion-based whole image classification using reduced feature set 65

7. Best confusion matrix results for fusion-based whole image classification using all 27 features 65

8. Fusion-based whole image normal versus cervical intraepithelial neoplasia and exact cervical intraepithelial neoplasia discrimination rates using all 27 features (F1-F27) with expert training-testing labels of RZ-SF and SF-RZ 66

9. Fusion-based whole image normal versus cervical intraepithelial neoplasia and exact cervical intraepithelial neoplasia discrimination rates using reduced features with training-testing labels of RZ-SF and SF-RZ 66

10. Confusion matrix classification baseline obtained from pathologist ground truth labels 68

11. Summary of best classification accuracies: Current study versus previous research versus current 70

PAPER III

1. Features Advances to Automatically Find Images for Application to Clinical Decision Support 82

2. Algorithm for generating HSV histogram. 85

3. Average test results using the nearest neighbor classifier for feature Groups 1-3 and benchmark features [3] for 20 training/test sets. 91

4. Multi-category test percentage correct results using nearest neighbor classifier for feature Groups 1-3 and benchmark features [3] for 20 training/test sets are presented with mean and standard deviation. 92

SECTION

1. INTRODUCTION

Data fusion can be defined as methods by which data from multiple sources can be combined to improve detection and decision accuracy. It has been of interest in areas such as military, remote sensing, finance, medical-diagnosis, clinical decision support, structural health monitoring and automated manufacturing. The fusion process strives to improve the signal-to-noise ratio of the sensor measurements by combining sensor data measuring similar physical phenomenon [17, 18, 19]. Besides improving the reliability of measured data, fusion algorithms also strive to combine complementary information from sensors and/or information sources measuring different physical phenomenon [20].

1.1. OVERVIEW OF DATA FUSION

Data fusion is a diverse collection of techniques which are used to "combine multiple sensor measurements and/or information from related sources to improve accuracy and draw more specific conclusions than would have been possible by using a single source" [20]. Data fusion is a multi-disciplinary research area that draws concepts from fields such as statistics, signal processing, computer science and artificial intelligence. Types of data fusion include low (raw-data) level-fusion, feature-level fusion and decision-level fusion. Raw-data fusion employs methods such as statistical detection and estimation techniques. Feature-level fusion is used to fuse descriptive features extracted from multiple sources measuring similar or dissimilar physical phenomena. This is done by combining the features into a single 1-D vector which could be used for discrimination purposes. Fusion of preliminary decisions/assessments made using data from single or multiple sources is

termed as decision-level data fusion [21, 22]. Data fusion techniques such as statistical methods, computational intelligence and machine learning, have been used by previous researchers for solving problems such as fault diagnosis, safety of complex systems, plant monitoring and monitoring in biomedicine [23, 24].

1.2. PROBLEM DESCRIPTION

In this dissertation, data fusion methods for uterine cancer intraepitheliual neoplasia (CIN) classification and a clinical decision support system are investigated. As for cervical cancer diagnosis, one of the methods for early cervical cancer diagnosis includes Pap test in which a colposcopy is used to visually inspect the cervix and microscopic interpretation of histology slides is performed by a pathologist when biopsied cervix tissue is available. Cervical intraepithelial neoplasia (CIN) is defined as the pre-malignant condition in which the atypical cells are limited in the epithelium region only [25, 26]. It has been observed that CIN shows progressively increasing atypical cells across the spectrum of the epithelium, from top to bottom, resulting in different CIN grades (CIN1, CIN2, CIN3) [25, 26, 27, 28, 22, 29]. It has been explored for CIN diagnosis in other studies [30, 28], that an epithelium with atypical cells could exhibit different CIN grades in different vertical segments of the epithelium [26]. Therefore to improve the classification accuracy, it is becoming necessary to analyze vertical segments of the epithelium and then fuse information from each of these segments to obtain the CIN grade of the whole epithelium region.

Biomedical information exists in different forms: as text and illustrations in journal articles and other documents, in "images" stored in databases. In the context of this work, an "image" includes not only biomedical images, but also illustrations, charts, graphs, and other visual material appearing in biomedical journals, and other relevant databases. Since medical images for a given modality (e.g. MRI, Histology or X-Ray) are very similar in nature, existing content-based image retrieval (CBIR) methods based only on the visual features of the whole image, such as texture and shape, may not be sufficient for accurate

retrieval of biomedical images [31]. In addition to using text (found in image captions and within the article) and visual features, retrieval of characters within images found in biomedical publications is an important task for obtaining complementary information for CBIR and analysis. Therefore, enhanced CBIR and analysis requires fusion of the various categories of information (image text, visual features and image captions).

1.3. SUMMARY OF CONTRIBUTIONS

This dissertation consists of three journal papers, as presented in the publication list. The unique contributions can be summarized as follows:

1.3.1. Improved Localized Classification of Cervical Intraepithelial Neoplasia (CIN) Diagnosis. The research presented for CIN diagnosis in this dissertation was developed and presented in [30]. Nuclei feature-based algorithm for analysis and classification of cervical caner diagnosis were developed. Based on the bounding box and medial axis algorithm in the previous research [32, 30], the whole epithelium region can be vertically divided into segments each of which is classified into one of the four CIN target grades. And nuclei features were developed using clustering methods with color information of pixels since the dark cells are presented with lower pixel values in most of the cases. Meanwhile, nuclei related features such as light area features are also developed with the similar algorithms by looking for the pixel clusters with relatively higher pixel values. It was observed during our experiments that the nuclei and nuclei-based features enhanced the original feature groups and provided a better classification of the CIN grades of the vertical segmented epithelium images. Based on the experimental results of vertical segments, a voting scheme is applied to fuse the multiple CIN grades into one final classification results for the whole epithelium region. And the fused classification performs better in exact classification, normal vs. CIN as well as leave-one-out scoring schemes, than to perform the classification upon the whole epithelium region with no fusion. The experiment showed the importance of data fusion at

the decision level for enhanced classification. The CIN diagnosis accuracy obtained using the vertical segments fusion method achieved 83%, which suggests the advantages of data fusion for enhanced classification accuracy.

1.3.2. Fusion Paradigm Development Based on Inter/Intra-pathologist CIN Labeling for Localized and Entire Image CIN Classification in Digitized Histology Images. As part of the CIN assessment process, acellular and atypical cell concentration features were computed from vertical segment partitions of the epithelium region to quantize the relative distribution of nuclei. Feature data was extracted from 610 individual segments from 61 images for epithelium classification into categories of Normal, CIN1, CIN2, and CIN3 [30, 33]. Individual vertical segment CIN classification accuracy improvement is obtained using the logistic regression classifier for an expanded data set of 118 histology images. The effects on classification were analyzed using the same pathologist labels for training and testing versus using one pathologist labels for training and the other for testing. Based on a leave-one-out approach for classifier training and testing, exact grade CIN accuracies of 81.29% and 88.98% were achieved for individual vertical segment and epithelium whole- image classification, respectively.

1.3.3. Advanced Features Development to Find Image for Application to Clinical Decision Support. Filtering through ever increasing sources of information to find relevant information for clinical decisions is a challenging task for clinicians. With the variety of items that can provide evidence to aid the decision-making process, illustration image analysis and classification has been used to characterize and distinguish specific image modalities; specifically, global, HSV histogram-based, and Gabor filter-based features are compared to histogram-based features for modality classification on a set of 12,056 images from 2004-2006 biomedical publication issues of Radiology and RadioGraphics that were manually annotated by modality (radiological, photo, etc.). Using a nearest-neighbor classifier, average modality discrimination results were obtained as high as 99.98% using

correlated features computed from Gabor filter spectral coefficients. These experimental results indicate that image features, particularly correlation-based features, can provide modality discrimination useful for clinical decision support applications.

PAPER

I. NUCLEI-BASED FEATURES FOR UTERINE CERVICAL CANCER HISTOLOGY IMAGE ANALYSIS WITH FUSION-BASED CLASSIFICATION

P. Guo[1] Koyel Banerjee[1] R.Joe Stanley[1] Rodney Long[2] Sameer Antani[2] George Thoma[2]

Rosemary Zuna[3] Shelliane R. Frazier[4] Randy H. Moss[1] William V.Stoecker[5]

[1]Department of Electrical & Computer Engineering

Missouri University of Science and Technology, Rolla

[2]Lister Hill National Center for Biomedical Communications,

National Library of Medicine, National Institutes of Health, DHHS, Bethesda, MD

[3]Department of Pathology,

University of Oklahoma Health Sciences Center, Oklahoma City, OK

[4]Surgical Pathology Department,

University of Missouri Hospitals and Clinics, Columbia, MO

[4]Stoecker & Associates, Rolla, MO, USA

Email: stanleyj@mst.edu

ABSTRACT

Cervical cancer, which has been affecting women worldwide as the second most common cancer, can be cured if detected early and treated well. Routinely, expert pathologists visually examine histology slides for cervix tissue abnormality assessment. In previous research, we investigated an automated, localized, fusion-based approach for classifying squamous epithelium into Normal, CIN1, CIN2, and CIN3 grades of cervical intraepithelial neoplasia (CIN) based on image analysis of 61 digitized histology images. This paper introduces novel acellular and atypical cell concentration features computed from vertical segment partitions

of the epithelium region within digitized histology images to quantize the relative increase in nuclei numbers as the CIN grade increases. Based on the CIN grade assessments from two expert pathologists, image-based epithelium classification is investigated with voting fusion of vertical segments using support vector machine and linear discriminant analysis approaches. Leave-one-out is used for the training and testing for CIN classification, achieving an exact grade labeling accuracy as high as 88.5%.

Keywords: Cervical cancer, cervical intraepithelial neoplasia (CIN), fusion-based classification, image processing, linear discriminant analysis (LDA), support vector machine (SVM).

1. INTRODUCTION

IN 2008, there were 529,000 new cases of invasive cervical cancer reported worldwide [30]. While the greatest impact of cervical cancer prevalence is in the developing world, invasive cervical cancer continues to be diagnosed in the U.S. each year. Detection of cervical cancer and its precursor lesion is accomplished through a Pap test, a colposcopy to visually inspect the cervix, and microscopic interpretation of histology slides by a pathologist when biopsied cervix tissue is available.

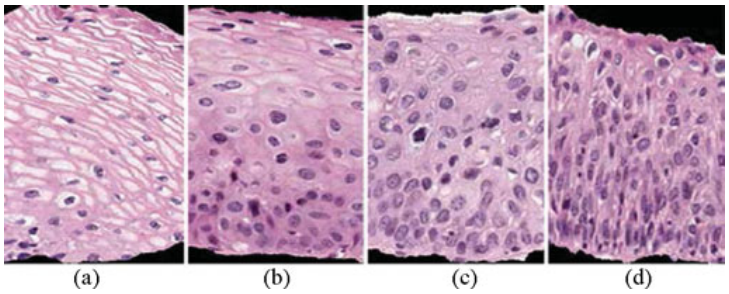

Figure 1. CIN grade label examples highlighting the increase of immature atypical cells from epithelium bottom to top with increasing CIN severity. (a) Normal. (b) CIN 1. (c) CIN 2. (d) CIN 3.

Microscopic evaluation of histology slides by a qualified pathologist has been used as a standard of diagnosis [13]. As a part of the pathologist diagnostic process, cervical intraepithelial neoplasia (CIN) is a premalignant condition for cervical cancer in which the atypical cells are identified in the epithelium by the visual inspection of histology slides [16]. As shown in Fig. 1, cervical biopsy diagnoses include Normal (that is, no CIN lesion), and three grades of CIN: CIN1, CIN2, and CIN3 [16, 10, 29]. CIN1 corresponds to mild dysplasia (abnormal change), whereas CIN2 and CIN3 are used to denote moderate dysplasia and severe dysplasia, respectively. Histologic criteria for CIN include increasing immaturity and cytologic atypia in the epithelium. As CIN increases in severity, the epithelium has been observed to show delayed maturation with an increase in immature atypical cells from bottom to top of the epithelium [10, 29, 26, 12]. As shown in Figure 1 atypical immature cells are seen mostly in the bottom third of the epithelium for CIN 1 [see Figure 1(b)]. For CIN2, the atypical immature cells appear in the bottom two thirds of the epithelium [see Figure 1(c)], and for CIN 3, atypical immature cells lie in the full thickness of the epithelium [see Figure 1(d)]. When these atypical cells extend beyond the epithelium, i.e., through the basement membrane and start to enter into the surrounding tissues and organs, it may indicate invasive cancer [16]. In addition to analyzing the progressively increasing quantity of atypical cells from bottom to top of the epithelium, identification of the nuclei atypia is also significant [16]. Nuclei atypia is a characteristic of nuclei enlargement, thereby resulting in different shapes and sizes of the nuclei present within the epithelium region. Visual assessment of this nuclei atypia may be difficult, due to the large number of nuclei present and the complex visual field, i.e., tissue heterogeneity. This may contribute to diagnostic grading repeatability problems and inter and intrapathologist variation [26, 12, 22]. Computer-assisted methods (digital pathology) have been explored for CIN diagnosis in other studies [5, 27, 8, 14], and provided the foundation for the work reported in [5]. These methods examined texture features [8], nuclei determination and Delaunay triangulation analysis [14], medial axis

determination [29], and localized CIN grade assessment [29]. A more detailed review of digital pathology techniques is presented in that paper [5]. Our research group previously investigated a localized fusion-based approach to classify the epithelium region into the different CIN grades, as determined by an expert pathologist [5]. We examined 66 features including texture, intensity shading, Delaunay triangle (DT) features (such as area and edge length), and weighted density distribution features, which yielded an exact CIN grade label classification result of 70.5% [5]. The goal of this paper, performed in collaboration with the National Library of Medicine, is to automatically classify 61 manually segmented cervical histology images into four different grades of CIN and to compare results with CIN grade determined by an expert pathologist. The research presented in this paper extends the study in [5] to the development of new image analysis and classification techniques for individual vertical segments to allow improved whole-image CIN grade determination. Specifically, we present new image analysis techniques to determine epithelium orientation and image analysis and to find and characterize acellular and nuclei regions within the epithelium. We also present comparative CIN grading classification analysis versus two expert pathologists CIN grading of the 61-image dataset. The order of the remaining sections of this paper is as follows. Section II presents the methods used in this paper. Section III describes the experiments performed. Section IV presents and analyzes the results obtained and a discussion. Section V provides the study conclusions.

2. METHODS

The images analyzed included 61 full-color digitized histology images of hematoxylin and eosinophil preparations of tissue sections of normal cervical tissue and three grades of cervical carcinoma in situ. An additional image, labeled as CIN1 by two experts (RZ and SF), was used for image processing algorithm parameter determination. The same experimental dataset was used in [5]. The entire classification process, as utilized in [5], of the segmented epithelium images was performed using the following five-step approach:

Step 1: Locate the medial axis of the segmented epithelium region.

Step 2: Divide the segmented image into ten vertical segments, orthogonal to the medial axis.

Step 3: Extract features from each of the vertical segments.

Step 4: Classify each of these segments into one of the CIN grades.

Step 5: Fuse the CIN grades from each vertical segment to obtain the CIN grade of the whole epithelium for image-based classification.

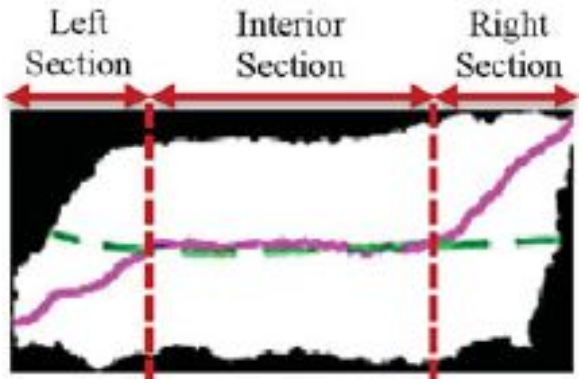

Figure 2. Example of incorrect medial axis determined using distance transform only (solid line). The desired medial axis is manually drawn and is overlaid on the image (dashed line). The left-hand, right-hand, and interior sections are labeled on the bounding box image to highlight the limitations of the distance transform algorithm.

The following sections present each step in detail.

2.1. Medial Axis Detection and Segments Creation. Medial axis determination used a distance-transform-based [20, 25] approach from [5]. The distance-transform-based approach from [5] had difficulties in finding the left- and right-hand end-axis portions of the epithelium axis in nearly rectangular and triangular regions. Figure. 2 shows an example of an incorrect medial axis estimation using a distance-transform-based approach (solid line) and the manually labeled desired medial axis (dashed line).

Accordingly, the algorithm from [5] used the bounding box of the epithelium to obtain a center line through the bounding box and intersecting the center line with the epithelium object. The resulting center line was divided into a left-hand segment (20%),

a right-hand segment (20%), and the interior segment (60%). These divisions of the epithelium can be observed in Figure. 2. The interior 60% portion of the distance-transform-based medial axis was retained as a part of the final medial axis. The left and right-hand cutoff points of the interior distance transform axis were determined as the closest Euclidean distance points from the distance transform axis to the center line points on the 20% left- and right-hand segments. As done in [5], the left- and right-hand cutoff points are projected to the median bounding box points for the remaining left-hand 20% and right-hand 20% portions of the axis. The projected left- and right-hand segments are connected with the interior distance transform axis to yield the final medial axis.

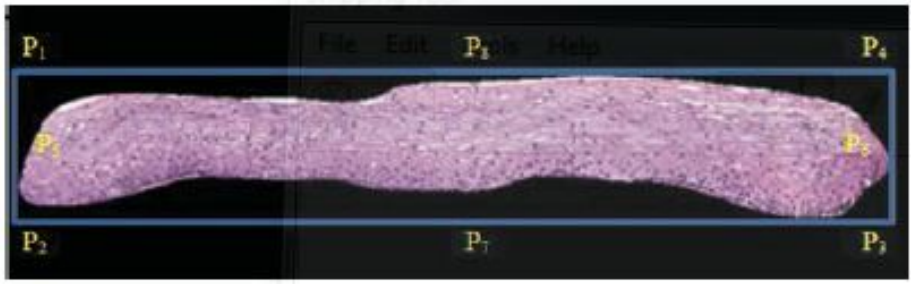

Figure 3. Bounding box of epithelium with control points labeled.

The epithelium's orientation was determined using a novel approach based on the bounding box and the final medial axis. Using the bounding box, a comparison was performed of the number of nuclei distributed over eight masks that are created from eight control points (P_1, P_2, P_3,..., P_8) at the corners and the midpoints of the bounding box edges (see Figure. 3).

The masks are used for computing the ratios of the number of detected nuclei to the areas of the masks. The control points used for determining the masks are shown as P_1 - P_8 in Figure. 3. For each control point combination, the number of nuclei is computed for each mask using the algorithm presented in Section II. Let n represent the set of the number of nuclei computed from masks 1 - 8, given as n = n_1, n_2, . . . , n_8, as designated in Figure. 4(a)-(d). The eccentricity, defined as the ratio of the fitted ellipse foci distance to the major

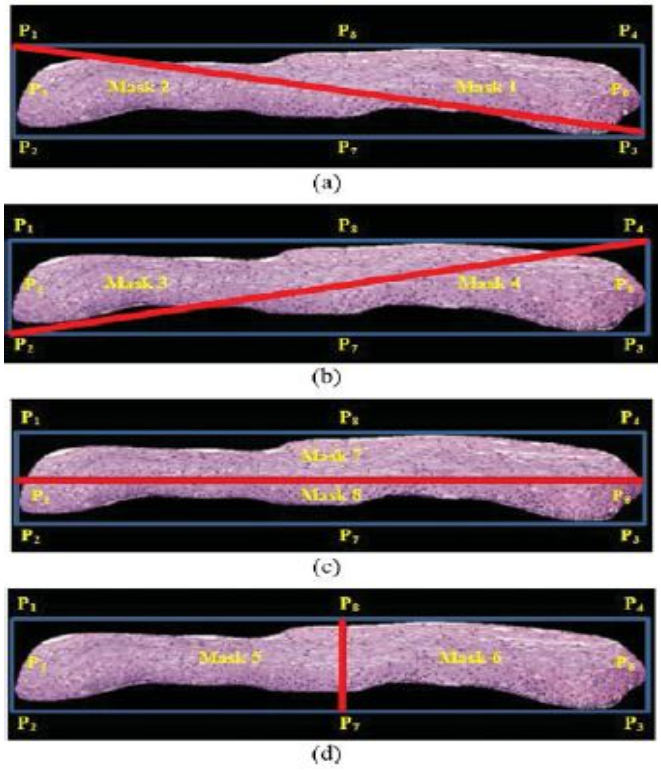

Figure 4. Bounding box partitioning with masks combinations shown based on control points from Fig. 3 as a part of epithelium orientation determination algorithm. (a) Mask 1 and Mask 2. (b) Mask 3 and Mask 4. (c) Mask 7 and Mask 8. (d) Mask 5 and Mask 6.

axis length as given in [19], is computed for the entire epithelium image mask, given as e, and for each mask image, denoted as e_i . Then, the eccentricity weighted nuclei ratios are calculated for each mask combination, given as v = v_{12}, v_{34}, v_{56}, v_{78}, v_{21}, v_{43}, v_{65}, v_{87}, where

$$v_{12} = \left(\frac{n_1}{max(n)}\right)\left(\frac{n_1}{n_2}\right)\left(\frac{e}{e_1}\right)$$

$$v_{34} = \left(\frac{n_3}{max(n)}\right)\left(\frac{n_3}{n_4}\right)\left(\frac{e}{e_3}\right) *$$

$$v_{56} = \left(\frac{n_5}{max(n)}\right)\left(\frac{n_5}{n_6}\right)\left(\frac{e}{e_5}\right)$$

$$v_{78} = \left(\frac{n_7}{max(n)}\right)\left(\frac{n_7}{n_8}\right)\left(\frac{e}{e_7}\right)$$

etc., and max(n) denotes the maximum area of the eight partitioned masks. The term $n_i/max(n)$ is used as a scale factor for normalizing the size of the epithelium region. The medial axis top/bottom orientation is determined as $v_{ij} = max_{ij}$ (v). The resulting medial axis is partitioned into ten segments of approximately equal length, perpendicular line slopes are estimated at the midpoints of each segment, and vertical lines are projected at the end points of each segment to generate ten vertical segments for analysis. The partitioning of the epithelium image into ten vertical segments was performed to facilitate localized CIN classifications within the epithelium that can be fused to provide an image-based CIN assessment, as done in [5]. Figure. 5 provides an example of the medial axis partitioning and the ten vertical segments obtained.

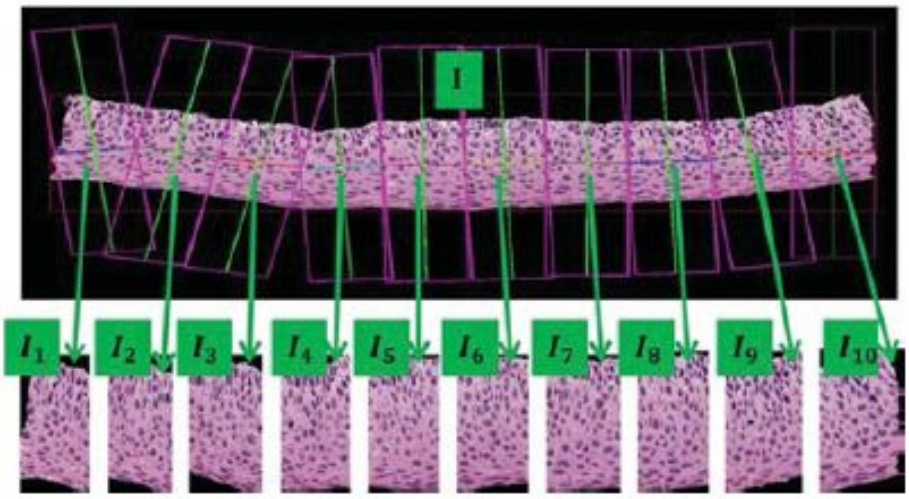

Figure 5. Epithelium image example with vertical segment images (I_1 ,I_2, I_3 ,..., I_{10}) determined from bounding boxes after dividing the medial axis into ten line segment approximations after medial axis computation.

2.2. Feature Extraction. Features are computed for each of the ten vertical segments of the whole image I_1 ,I_2 ,I_3, . . . , I_{10}. All the segments of one whole image are feature extracted in a sequence, from left to right I_1 to I_{10} (see Figure 5). In total, six different types of features were obtained in this study, including: 1) texture features (F1-F10) [5]; 2) cellularity features (F11-F13); 3) nuclei features (F14, F15); 4) acellular (light area) features (F16-F22); 5) combination features (F23, F24); and 6) advanced layer-by-layer triangle features (F25- F27). To give a brief introduction of the extracted features, Table I is presented showing the feature label and brief description in every row for each feature.

2.2.1. Texture and Cellular Features. The texture and color features were used in our previous work and are described in [5]. The use of color in histopathological image analysis is also described in [27] and [8]. For texture features, both first-order structural measures derived directly from the image segment and second-order statistical methods based on the gray-level cooccurrence matrix (GLCM) [29, 3], were employed. A gray-scale luminance version of the image was created in order to compute the statistics of energy, correlation, contrast, and uniformity of the segmented region; these statistics are then used to generate features (F1-F10) shown in Table I. The texture features include contrast (F1), energy (F2), correlation (F3), and uniformity (F4) of the segmented region, combined with the same statistics (contrast, energy, and correlation) generated from the GLCM of the segment (F5-F10, see Table 1).

The luminance images showed regions with three different intensities, marked as light, medium, and dark areas within each single-segmented luminance image, as shown in Figure 6 of normal cervical histology. The light areas correspond to acellular areas; the medium areas correspond to cytoplasm; and the dark areas correspond to nuclei.

Cluster centers are found from the luminance image using K-means clustering [9] for three different regions (K = 3) denoted as clustLight, clustMedium, and clustDark for the light, medium, and dark cluster centers, respectively. Then, the ratios are calculated based on (1)-(3) [5]

Table 1. Feature description

Label	Description
F1	Contrast of segment: Intensity contrast between a pixel and its neighbor over the segment image.
F2	Energy of segment: Squared sum of pixel values in the segment image.
F3	Correlation of segment: How correlated a pixel is to neighbors over the segment image.
F4	Segment homogeneity: Closeness of the distribution of pixels in the segment image to the diagonal elements.
F5, F6	Contrast of GLCM: Local variation in GLCM in horizontal and vertical directions.
F7, F8	Correlation of GLCM: Joint probability occurrence (periodicity) of elements in the segment image in the horizontal and vertical directions.
F9, F10	Energy of GLCM: Sum of squared elements in the GLCM in horizontal and vertical directions.
F11	Acellular ratio: Proportion of object regions within segment image withlight pixels (acellular).
F12	Cytoplasm ratio: Proportion of object regions within segment image withmedium pixels (cytoplasm).
F13	Nuclei ratio: Proportion of object regions within segment image with dark pixels (nuclei).
F14	Average nucleus area: Ratio of total nuclei area over total number of nuclei.
F15	Background to nuclei area ratio: Ratio of total background area to total nuclei area.
F16	Intensity ratio: Ratio of average light area image intensity to background intensity.
F17	Ratio R: Ratio of average light area red to background red.
F18	Ratio G: Ratio of average light area green to background green.
F19	Ratio B: Ratio of average light area blue to background blue.
F20	Luminance ratio: Ratio of average light area luminance to background luminance.
F21	Ratio light area: Ratio of light area to total area.
F22	Light area to background area ratio: Ratio of total light area to background area.
F23	Ratio acellular number to nuclei number: Ratio of number of light areasto number of nuclei.
F24	Ratio acellular area to nuclei area: Ratio of total light area to total nuclei area.
F25	Triangles in top layer: Number of triangles in top layer.
F26	Triangles in mid layer: Number of triangles in middle layer.
F27	Triangles in bottom layer: Number of triangles in bottom layer.

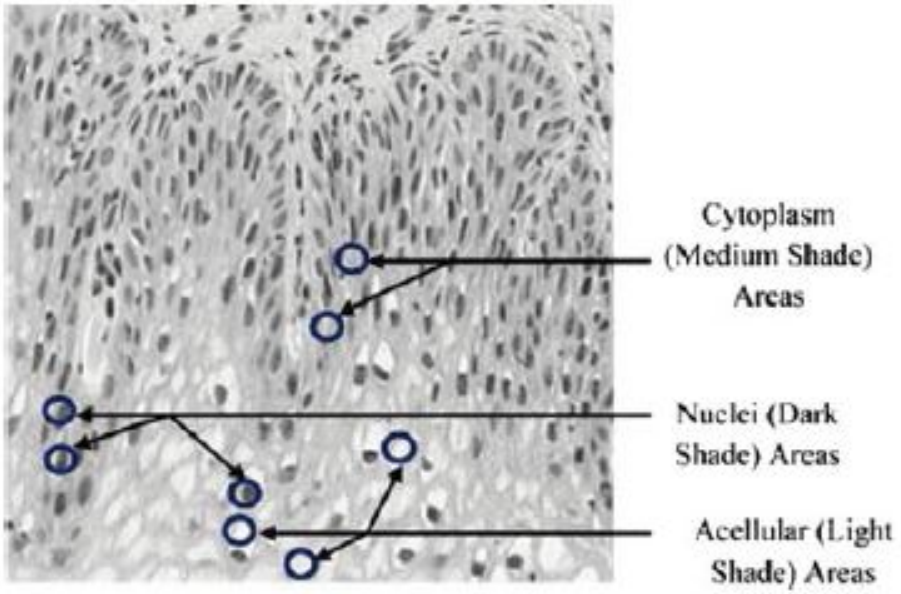

Figure 6. Sample shading representatives within epithelium image used for determining cellular features.

$$\text{Accellular ratio} = \frac{\text{number of pixels in light cluster}}{\text{Total pixels in epithelium region}} \tag{1}$$

$$\text{Cytoplasm ratio} = \frac{\text{number of pixels in medium cluster}}{\text{Total pixels in epithelium region}} \tag{2}$$

$$\text{Nuclei ratio} = \frac{\text{number of pixels in dark cluster}}{\text{Total pixels in epithelium region}} \tag{3}$$

where Acellular ratio (F11), Cytoplasm ratio (F12), and Nuclei ratio (F13) represent the cellular features in Table I, and numLight, numMedium, and numDark represent the number of pixels that were assigned to the clusters of light, medium, and dark, respectively. These features correspond to intensity shading features developed in [5].

2.2.2. Nuclei Features. The dark shading color feature discussed above corresponds to nuclei, which appear within epithelial cells in various shapes and sizes. Nuclei tend to increase in both number and size as the CIN level increases. This linkage between

the nuclei characteristics and CIN levels motivates our development of algorithms for nuclei detection feature extraction. In this paper, the algorithms of nuclei detection and nuclei feature extraction are developed to obtain features to facilitate CIN classification. Specifically, we carry out the following steps:

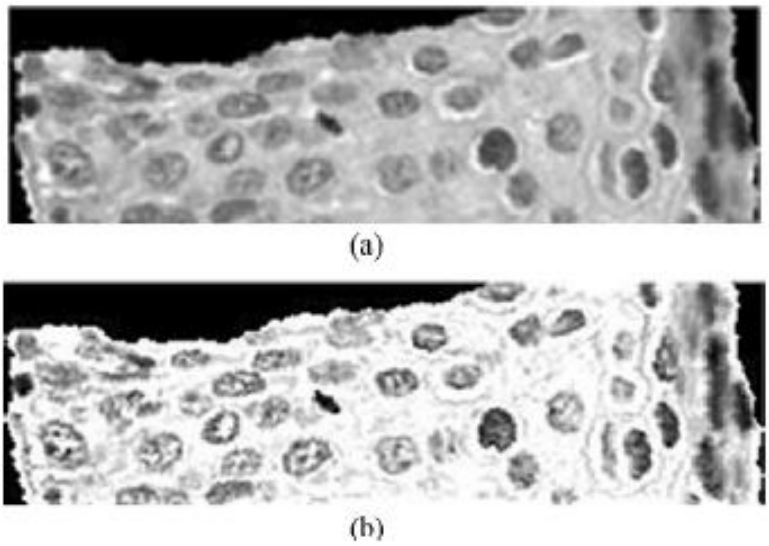

Figure 7. Example of image preprocessing. (a) Original luminance image I. (b) Sharpened image $I_{sharpen}$ obtained after average filtering of I.

1. nuclei feature preprocessing: average filter, image sharpening, and histogram equalization;

2. nuclei detection: clustering, hole filling, small-area elimination, etc;

3. nuclei feature extraction.

In preprocessing, vertical image segments are processed individually. After converting the segment into a gray-scale image I, an averaging filter is applied as in (4), where * denotes convolution

$$A = \frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} * I \tag{4}$$

After the average-filtered image is obtained, an image sharpening method is used to emphasis the dark shading part, which is expressed as (5), following the methods in [7]

$$I_{Sharpen} = kI - A \tag{5}$$

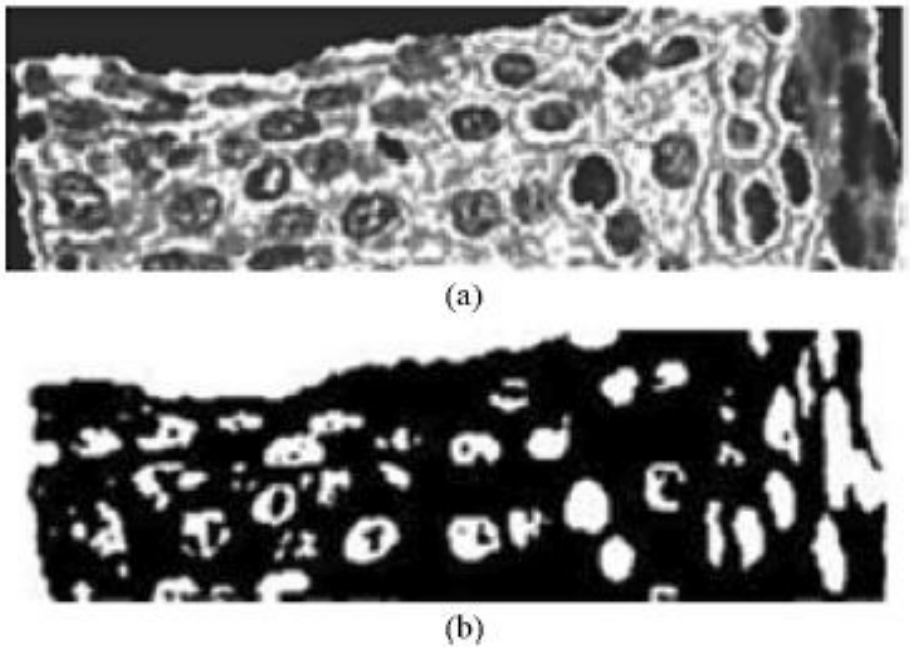

Figure 8. Example of image processing steps to obtain nuclei cluster pixels from K-means algorithm from histogram equalized image. (a) Histogram equalized image determined from Figure. 7(b). (b) Mask image obtained from K-means algorithm with pixels closest to nuclei cluster.

where $I_{sharpen}$ is the sharpened image, and k is an empirically determined constant of 2.25. The average-filtered image A and the sharpened image $I_{sharpen}$ are shown in Figure. 8. In the final preprocessing step, we apply histogram equalization using the MATLAB function $histeq$ to the sharpened image ($I_{sharpen}$) (in particular, to enhance details of the nuclei atypia).

The nuclei detection algorithm is described as follows using the equalized histogram image as the input.

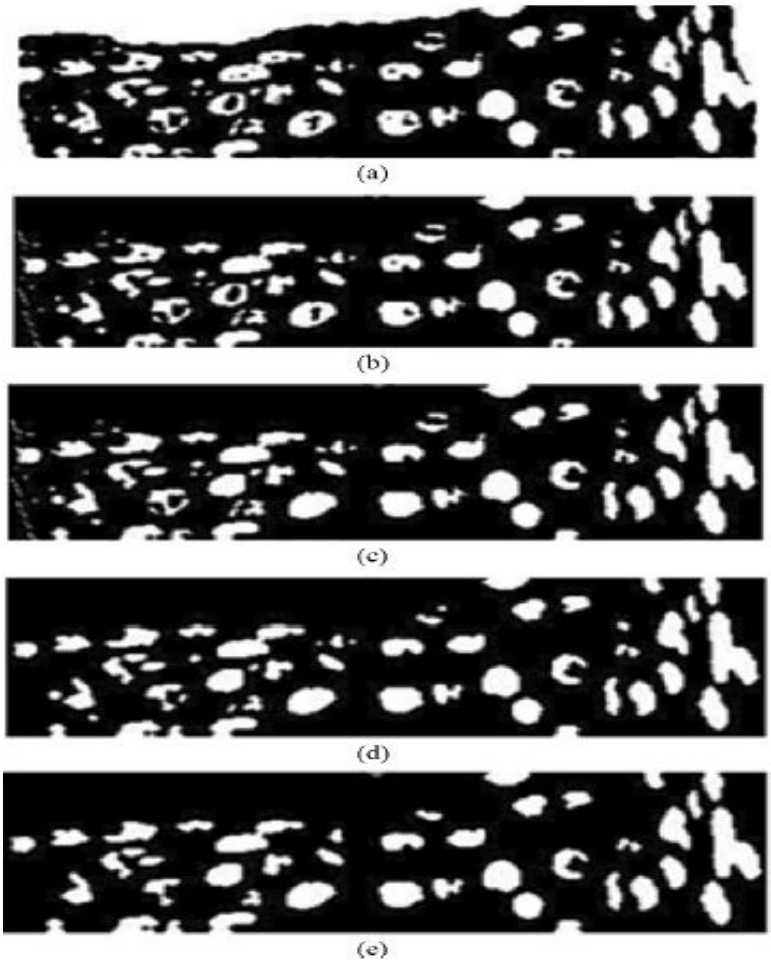

Figure 9. Image examples of nuclei detection algorithm. (a) Image with preliminary nuclei objects obtained from clustering [step 1-Figure 8(c)]. (b) Image closing to connect nuclei objects (step 2). (c) Image with hole filling to produce nuclei objects (step 3). (d) Image opening to separate nuclei objects (step 4). (e) Image with nonnuclei (small) objects eliminated (step 5).

1. Step 1: Cluster the histogram-equalized image into clusters of background (darkest), nuclei, and darker and lighter (lightest) epithelium regions using the K-means algorithm (K = 4). Generate a mask image containing the pixels closest to the nuclei cluster (second darkest).

2. Step 2: Use the MATLAB function *imclose* with a circular structuring element of radius 4 to perform morphological closing on the nuclei mask image.

3. Step 3: Fill the holes in the image from Step 2 with MATLAB's $imfill$ function for this process.

4. Step 4: Use the MATLAB'S *imopen* to perform morphological opening with a circular structuring element of radius 4 on the image from Step 3.

5. Step 5: Eliminate small area noise objects (nonnuclei objects) within the epithelium region of interest from the mask in Step 4, with the area opening operation using the MATLAB function *bwareaopen*.

Figure 8 shows an example of a sharpened image before and after histogram equalization, which is input to the nuclei detection algorithm, and the resulting mask image with pixels closest to the nuclei cluster from the K-means algorithm in Step 1. The nuclei detection algorithm steps 2-5 are illustrated in Fig. 9. The nuclei features are calculated as follows: With the detected nuclei shown as white objects in the final binary images [see Figure. 9(e)], the nuclei features are calculated as

$$Average\ nucleus\ area = \frac{nuclei\ area\ total}{number\ of\ nuclei} \tag{6}$$

$$Ratio\ of\ background\ to\ nucleus\ area = \frac{total\ non-nuclei\ area}{total\ nuclei\ area} \tag{7}$$

where Average nucleus area (F14) and Ratio of background to nuclei area (F15) represent ratios obtained from the final nuclei images as shown in Figure. 9(e). In (6) and (7), NucleiArea represents the total area for all nuclei detected (all white pixels); $NucleiNumber$ indicates the total number of white regions [number of objects in Figure 9(e)]; $AverageNucleusArea$ is the ratio of nuclei area to the nuclei number, which tends to increase with higher CIN grade; $NonNucleiArea$ area represents the total number of pixels in the black nonnucleus region within the epithelium region [black pixels within epithelium

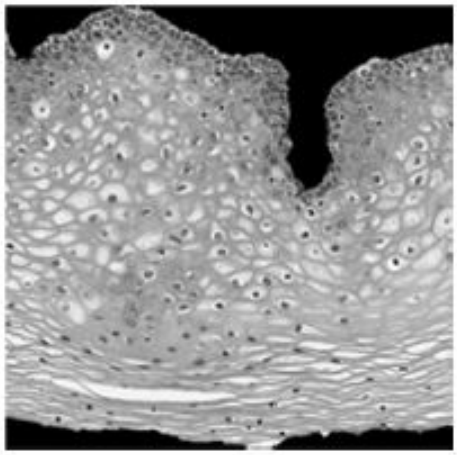

Figure 10. Example L* image for light area detection.

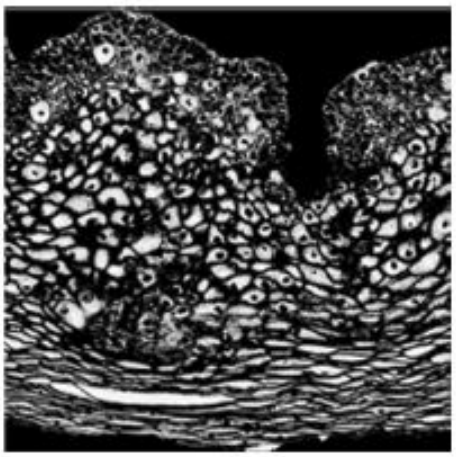

Figure 11. Adaptive histogram equalized image of Fig. 10.

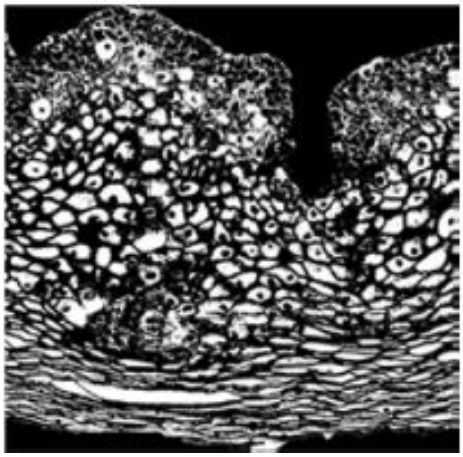

Figure 12. Thresholded image of Fig. 11.

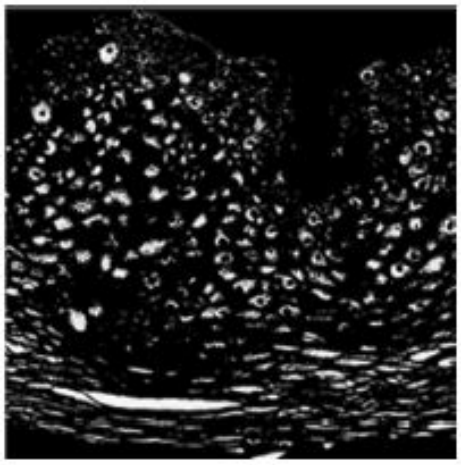

Figure 13. Example image of light area clusters after K-means clustering.

in Fig 9(e)]. *RatioBackgroundNucleusArea* denotes the ratio of the nonnuclei area to nuclei area.We expect larger values of *AverageNucleusArea* to correspond to increasing CIN grade, and *RatioBackgroundNucleusArea* to decrease with increasing CIN grade.

2.2.3. Acellular Features. Extracting the light area regions, described previously as "light shading," is challenging due to the color and intensity variations in the epithelium images. We evaluated each of the L*, a*, and b* planes of CIELAB color space for characterizing the light areas, and determined empirically that L* provides the best visual results. The following outlines the methods we used to segment the histology images.

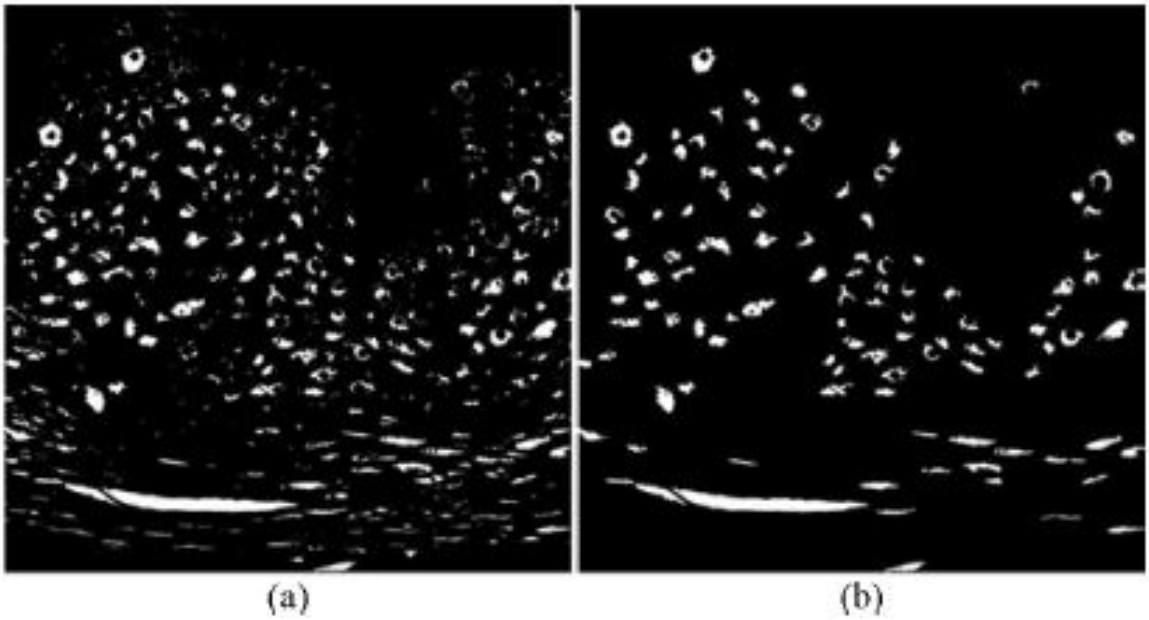

Figure 14. Example morphological dilation and final light area mask. (a) Morphological dilation and erosion process after K-means clustering. (b) Final light area mask, after eliminating regions with areas smaller than 100 pixels.

1. Step 1: Convert the original image from RGB color space to L*a*b* color space, then select the luminance component L* (see Fig. 10).

2. Step 2: Perform adaptive histogram equalization on the image from step 1 using MATLAB'S *adapthisteq*. *adapthisteq* operates on small regions (tiles) [2] for contrast enhancement so that that the histogram of the output region matches a specified histogram and combines neighboring tiles using bilinear interpolation to eliminate artificially induced boundaries (see Fig. 11).

3. Step 3: After the image has been contrast adjusted, the image is *binarized* by applying an empirically determined threshold of 0.6. This step is intended to eliminate the dark nuclei regions and to retain the lighter nuclei and epithelium along with the light areas (see Fig. 12).

4. Step 4: Segment the light areas using the K-means algorithm based on [5], with K equal to 4. The K-means algorithm input is the histogram-equalized image from Step 2 multiplied by the binary thresholded image from Step 3. A light area clustering example is given in Fig. 13.

5. Step 5: Remove all objects having an area less than 100 pixels from the image, determined empirically, using the MATLAB function regionprops [2]. A morphological closing is performed with a disk structure element of radius 2. An example result is shown in Fig. 14.

Using the light area mask, the acellular features (from Table I) are computed and are given as follows:

$$Intensity\ ratio = \frac{Light\ Area\ Intensity}{Background\ Intensity} \tag{8}$$

$$Ratio\ R = \frac{Light\ Area\ Red}{Background\ Red} \tag{9}$$

$$Ratio\ \mathrm{G} = \frac{Light\ Area\ Green}{Background\ Green} \tag{10}$$

$$Ratio\ B = \frac{Light\ Area\ Blue}{Background\ Blue} \tag{11}$$

$$Luminance\ ratio = \frac{LightArealuminance}{Backgroundluminance} \tag{12}$$

$$Ratio\ acellular\ number\ to\ light\ number = \frac{LightNumber}{NucleiNumber} \tag{13}$$

$$Ratio\ acellular\ area\ to\ light\ area = \frac{LightArea}{NucleiArea} \tag{14}$$

where $SegmentArea$ gives the epithelium area within the vertical segment; $LightArea$ denotes the area of all light area regions; $LightNumber$ corresponds to the number of light areas; $BackgroundArea$ represents the total number of nonnuclei and nonlight area pixels inside the epithelium within the vertical segment (i.e., background area); $LightAreaIntensity$, $LightAreaRed$, $LightAreaGreen$, $LightAreaBlue$, and $LightAreaLuminance$ are the average intensity, red, green, blue, and luminance values, respectively, of the light areas within the epithelium of the vertical segment; $BackgroundIntensity$, $BackgroundRed$, $BackgroundGreen$, $BackgroundBlue$, and $BackgroundLuminance$ are the average intensity, red, green, blue, and luminance values, respectively, of the nonnuclei and nonlight area pixels within the epithelium of the vertical segment. 4) Combination features: After both the nuclei features and the acellular features were extracted, two new features were calculated with the intent to capture the relative increase in nuclei numbers as CIN grade increases. These features are the ratio of the acellular number to the nuclei number (F23), and the ratio of the acellular area to the total nuclei area (F24). The equations for calculating the combination features are presented as follows:

$$Ratio\ acellular\ number\ to\ nuclei\ number = \frac{Light\ Number}{Nuclei\ Number} \tag{15}$$

$$Ratio\ acellular\ area\ to\ nuclei\ area = \frac{Light\ Area}{Nuclei\ Area} \tag{16}$$

where $LightNumber$ and $NucleiNumber$ represent the total number of light area and nuclei objects, respectively, as found in Sections II (see Fig. 6).

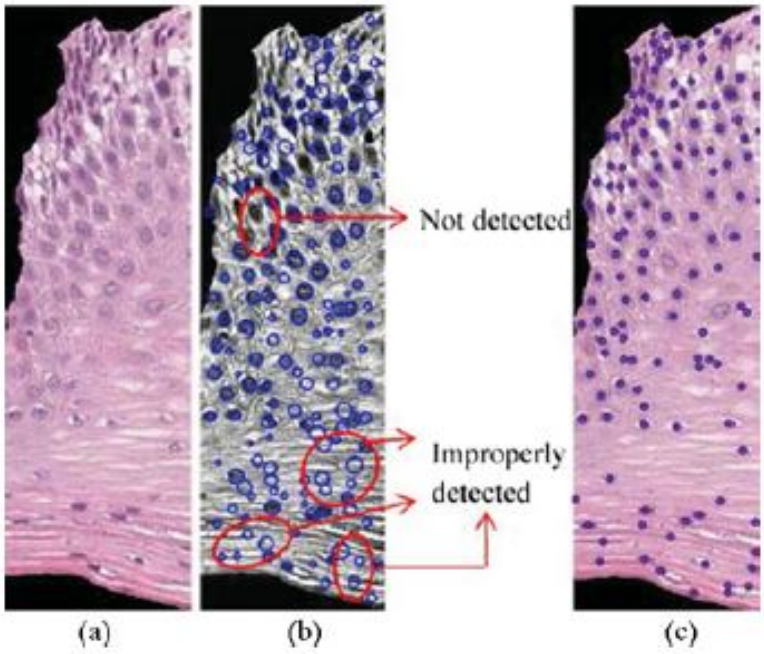

Figure 15. Example of nuclei detection comparison between the circular-Hough method and the method presented in this paper. (a) Original vertical segment. (b) Example of circular-Hough method; note the nuclei misses and false detections. (c) Nuclei detected using the algorithm from Section II.

2.2.4. Triangle Features. In previous research, triangle features have been investigated based on the circular-Hough Transform (CHT) [3] to detect nuclei for use as vertices in DT formulation [24] to obtain the triangles [5, 27]. Features were computed which included triangle area and edges length, and simple statistics (means, standard deviations) of these quantities were also included as features. In applying the CHT to our experimental dataset, we observed that for some images, this method sometimes fails to locate noncircular irregularly-shaped nuclei; on the other hand, this method does (incorrectly) detect some nonnuclei regions as nuclei, which leads to incorrect vertices being input to the (DT) method, thus degrading the triangle features calculated in downstream processing. To overcome the shortcomings of the CHT for nuclei detection, we use the centroids of

the nuclei detected based on the method presented in Section II. An example comparison between the previous circular-Hough method and the method in this paper is presented in Figure. 15. Circles indicate the locations of detected nuclei.

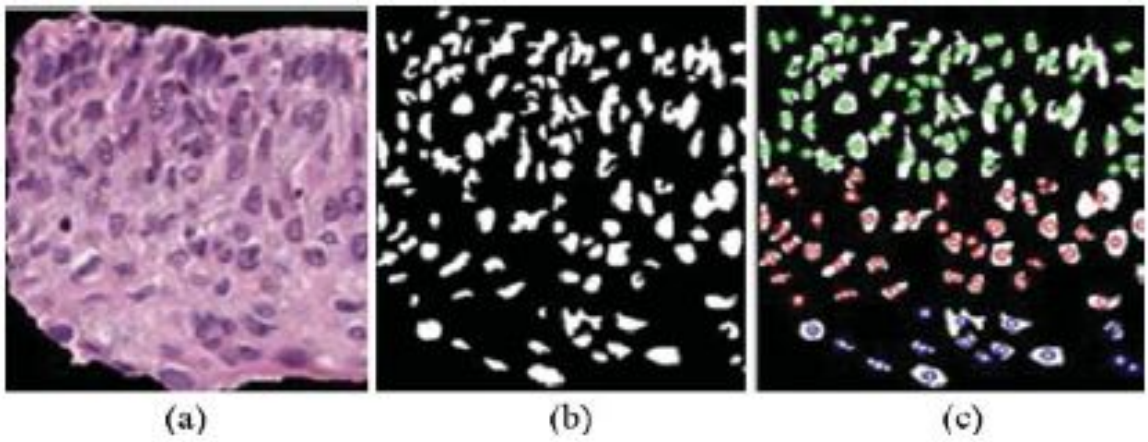

Figure 16. Distribution of nuclei centroids as vertices for DTs in bottom layer (green), mid layer (red), and top layer (blue).

In this paper, we use the DT method, but restrict the geometrical regions it can act upon, as follows. Before forming the DTs with the vertices provided by the nuclei detection results from Section II, the vertical segment being processed is subdivided into three vertical layers as illustrated in Figure. 16. The aim is to associate the presence of increasing nuclei throughout the epithelium with increasing CIN grades, namely: abnormality of the bottom third of the epithelium roughly corresponds to CIN1; abnormality of the bottom two-thirds, to CIN2; and abnormality of all three layers, to CIN3. We refer to these layers as bottom, mid, and top. (See Figure 16, the green circles stand for the top layer vertices, red for mid layer, and blue for bottom.)

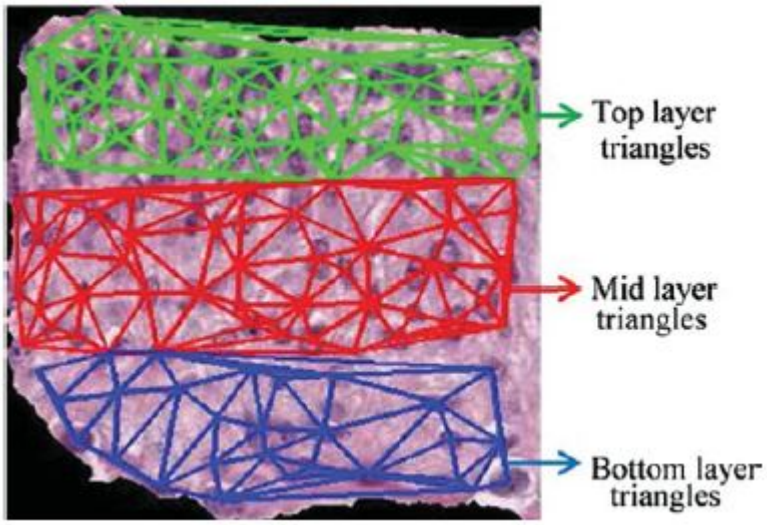

Figure 17. DTs in bottom layer (green lines), mid layer (red lines), and top layer (blue lines).

After locating the vertices for DT, the DT algorithm iteratively selects point triples to become vertices of each new triangle created. Delaunay triangulation exhibits the property that no point lies within the circles that are formed by joining the vertices of the triangles [29]. As shown in Fig. 17, all the triangles in three layers formed using DT are unique and do not contain any points within the triangles. The features are obtained according to the triangles in three different layers, including the number of triangles in top layer (F25), number of triangles in middle layer (F26), and number of triangles in bottom layer (F27).

3. EXPERIMENT PERFORMED

We carried out three sets of experiments, which are described in this section. The experimental dataset consisted of 61 digitized histology images, which were CIN labeled by two experts (RZ and SF) (RZ: 16 Normal, 13 CIN1, 14 CIN2, and 18 CIN3; SF: 14 Normal, 14 CIN1, 17 CIN2, and 16 CIN3).

3.1. Fusion-Based CIN Grade Classification of Vertical Segment Images. For the first set of experiments, all the features extracted from the vertical segment images were used as inputs to train the SVM/LDA classifier. The LIBSVM [4] implementation of the

SVM [15] and LDA [6] classifiers were utilized in this study. The SVM implementation uses a linear kernel and the four weights were the fractions of the images in each CIN class to the entire image set (fraction of the image set that is Normal, fraction of the image set that is CIN1, etc.).

Individual features were normalized by subtracting the mean training set feature value and dividing by the standard deviation training set feature value. Due to the limited size of the image set, a leave-one-image-out approach was investigated for classifier training and testing. For this approach, the classifier is trained based on the individual vertical segment feature vectors for all but the left-out epithelium image (used as the test image). For classifier training, the expert truthed CIN grade for each image was assigned to the ten vertical segments for that image. For the left-out test image, each vertical segment was classified into one of the CIN grades using the SVM/LDA classifier.

Then, the CIN grades of the vertical segment images were fused to obtain the CIN grade of the entire test epithelium image (see Figure. 5). The fusion of the CIN grades of the vertical segment images was completed using a voting scheme. The CIN grade of the test epithelium image was assigned to the most frequently occurring class assignment for the ten vertical segments. In the case of a tie among the most frequently occurring class assignments for the vertical segments, the test image is assigned to the higher CIN class. For example, if there is a tie between CIN1 and CIN2, then the image is designated as CIN2. The leave-one-image-out training/testing approach was performed separately for each expert's CIN labeling of the experimental dataset. For evaluation of epithelium classification, three scoring schemes were implemented.

Scheme 1 (Exact Class Label): The first approach is exact classification meaning that if the class label automatically assigned to the test image is the same as the expert class label, then the image is considered to be correctly labeled. Otherwise, the image is considered to be incorrectly labeled.

Scheme 2 (Off-by-One Class Label): The second scoring approach is an Off-by-One classification scheme, known as "windowed class label" in previous research [5]. If the predicted CIN grade level for the epithelium image is only one grade off as compared to the expert class label, the classification result is considered correct. For example, if the expert class label CIN2 was classified as CIN1 or CIN 3, the result would be considered correct. If expert class label CIN1 was classified as CIN3, the result would be considered incorrect.

Scheme 3 (Normal Versus CIN): For the third scoring scheme, the classification result would be considered incorrect when a Normal grade was classified as any CIN grade and vice versa.

3.1.1. Classification of the Whole Epithelium. For the second set of experiments, features were extracted from the whole epithelium image following the steps shown in Fig. 18, which also gives the comparison between the whole epithelium image classification and fusion-based classification over vertical segments (see Section III).

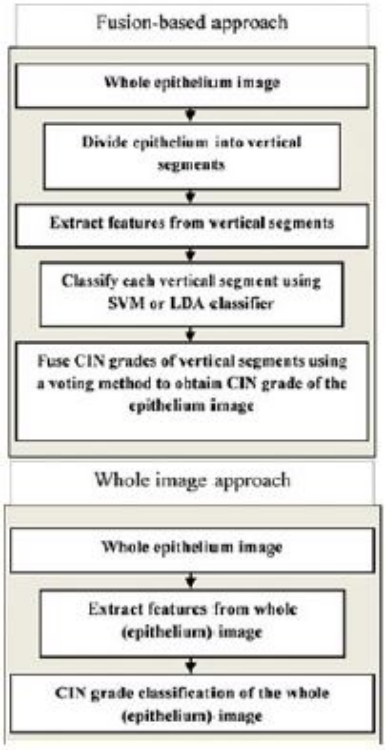

Figure 18. Fusion-based approach versus whole image approach.

The whole-epithelium image classification in this section is done without generating any of the individual vertical segment images (see Figure 19 as an example for nuclei feature detection over the whole image). The experiment was investigated to compare the performance of the fusion-based epithelium classification (see Section III) to the performance obtained by classifying the epithelium image as a whole. Features extracted from the whole image were used as inputs to the SVM/LDA classifier using the same leave-one-image out approach. The same scoring schemes as presented in Section III were used to evaluate the performance of the whole epithelium classification. Again, the leave-one-image-out training/testing approach was performed separately for each expert's CIN labeling of the experimental dataset.

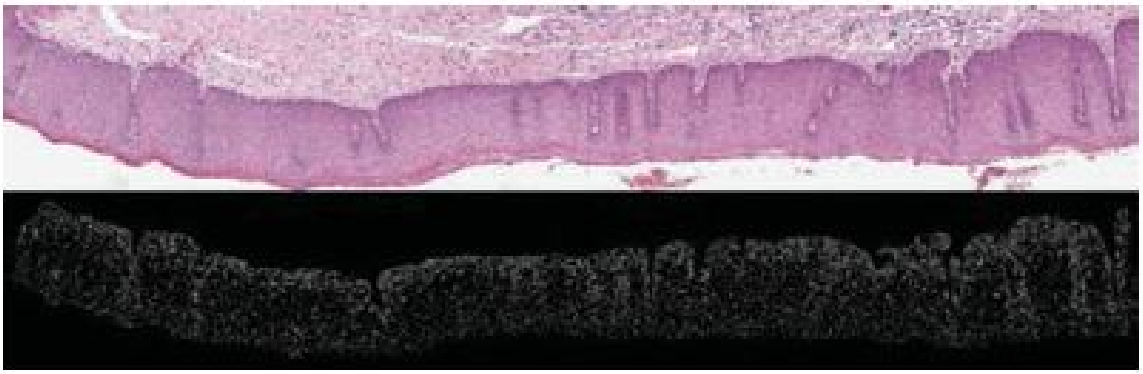

Figure 19. Example image of nuclei detection over whole image without creating vertical segments; the top image is the original epithelium image; bottom is the nuclei mask of this image.

3.1.2. Feature Evaluation and Selection. For feature evaluation and selection, a SAS implementation of multinomial logistic regression (MLR) [23, 17, 11, 9] and a Weka attribute information gain evaluator were employed. For SAS analysis, MLR is used for modeling nominal outcome variables, where the log odds of the outcomes are modeled as a linear combination of the predictor variables [23, 17, 11, 1]. The p-values obtained from the MLR output are used as a criterion for selecting features with p-values less than an appropriate alpha (α) value [23, 17, 11, 1]. For Weka analysis, the algorithm ranks the

Table 2. Confusion matrix results for fusion-based classification using all 27 features (F1-F27) for SVM and LDA classifiers for both experts

	Expert RZ: SVM/LDA			
	Normal (16)	CIN1 (13)	CIN2 (14)	CIN3 (18)
Normal	14/14	0/0	0/0	0/0
CIN1	2/2	12/11	0/0	0/0
CIN2	0/0	1/1	12/14	3/3
CIN3	0/0	0/1	2/0	15/15
	Expert SF: SVM/LDA			
	Normal (14)	CIN1 (14)	CIN2 (17)	CIN3 (16)
Normal	10/10	2/3	0/0	0/0
CIN1	4/3	9/9	1/1	0/0
CIN2	0/0	3/1	16/16	1/1
CIN3	0/1	0/1	0/0	15/15

features by a parameter called "attributes information gain ratio (AIGR)," where the higher the ratio, the more significant the feature will be for the classification results. For both methods, the automatically generated labels of the vertical segmentations and the feature data are given as input.

4. EXPERIMENTAL RESULTS AND ANALYSIS

4.1. Experimental Results. We obtained the vertical segment image classifications (CIN grading) using the SVM/LDA classifier with a leave-one-image-out approach based on all the 27 features generated. Then, the vertical segment classifications were fused using a voting scheme to obtain the CIN grade of the epithelium image. We evaluated the performance of these epithelium image classifications using the three approaches presented in Section III. Table II shows the confusion matrices for the classification results obtained using the fusion-based approach, for the SVM and LDA classifiers, respectively, for both experts (RZ and SF).

Table 3. Confusion matrix results for whole image classification using all 27 features (F1-F27) for SVM and LDA classifiers for both experts

	Expert RZ: SVM/LDA			
	Normal (16)	CIN1 (13)	CIN2 (14)	CIN3 (18)
Normal	15/9	0/4	0/1	0/0
CIN1	1/5	8/5	3/2	3/1
CIN2	0/1	2/3	8/8	5/8
CIN3	0/1	3/1	3/3	10/9
	Expert SF: SVM/LDA			
	Normal (14)	CIN1 (14)	CIN2 (17)	CIN3 (16)
Normal	9/10	2/3	0/0	0/0
CIN1	5/3	6/8	2/3	1/1
CIN2	0/1	5/2	11/8	8/8
CIN3	0/0	1/1	4/6	7/7

In the following, we provide summary comments for these Table 2 results, and compare them with the previous results published in [5], which used the RZ expert CIN labeling of the image set. 1) For the Exact Class Label, we obtained an accuracy of 86.9%/82.0% (RZ/SF) using the SVM classifier and 88.5%/82.0% (RZ/SF) using the LDA classifier, (previous [5]: 62.3% LDA). 2) For the Normal Versus CIN scoring scheme, SVM classifier accuracy was 96.7%/90.2% (RZ/SF) and LDA classifier 96.7%/90.2% (RZ/SF) (previous [5]: 88.5% LDA). 3) For the Off-by-One class scoring scheme, SVM had an accuracy of 100%/100% (RZ/SF) and LDA, 98.4%/96.7% (RZ/SF) (previous [5]: 96.7%).

In the order to evaluate the performance of the fusion-based approach for epithelium classification, we also carried out classification using the entire epithelium image. For the whole image classification, we again used the SVM and LDA classifiers. Table 3 shows the whole image classification results for both experts.

From Table 3, the Exact Class Label scoring scheme provided an accuracy of 67.2%/54.1% (RZ/SF) and 50.8%/54.1% (RZ/SF) using the SVM and LDA classifiers, respectively. The Normal versus CIN scoring scheme yielded an accuracy of 98.4%/88.6%

Table 4. Confusion matrix results for fusion-based classification using reduced features for SVM and LDA classifiers for both experts

	Expert RZ: SVM/LDA			
	Normal (16)	CIN1 (13)	CIN2 (14)	CIN3 (18)
Normal	14/14	0/0	0/0	0/0
CIN1	2/2	10/12	0/0	0/0
CIN2	0/0	2/1	12/13	3/1
CIN3	0/0	1/0	2/1	15/17
	Expert SF: SVM/LDA			
	Normal (14)	CIN1 (14)	CIN2 (17)	CIN3 (16)
Normal	11/10	3/2	0/0	0/0
CIN1	3/3	9/11	1/1	0/0
CIN2	0/1	2/1	16/16	1/1
CIN3	0/0	0/0	0/0	15/15

(RZ/SF) and 80.3%/88.6% (RZ/SF) for the SVM and LDA classifiers, respectively. The Off-by-One scoring scheme obtained an accuracy of 90.2%/96.7% (RZ/SF) and 93.4%/95.1% for the SVM and LDA classifiers, respectively. The corresponding accuracy figures from the previous research [9] for the LDA classifier are given in the following. Exact Class Label scoring: 39.3%; Normal versus CIN scoring: 78.7%; Off-by-One scoring (called "windowed class" in [5]): 77.0%.

For feature evaluation and selection experiments, all 27 features extracted from the individual vertical segments were used as inputs to the SAS MLR algorithm. We used α = 0.05 as the threshold to determine statistically significant features. The overall 27 features with p-values are presented in Table 6 (see appendix). From Table 5, features with a p-value smaller than 0.05 are considered statistically significant features.

In addition, all 27 features and the truth labels were used as input for the Weka information gain evaluation algorithm [9]. The algorithm ranks the features by an AIGR which ranges from 0 to 1. The larger the ratio is, the more likely that the feature is considered by the algorithm. The 27 features and corresponding AIGR values are shown in Table 5 .

Table 5. Features with corresponding p-values and AIGR

Feature	p-value	AIGR	Feature	p-value	AIGR
F1	0.0013	0.226	F15	0.1101	0.505
F2	>0.05	0.21	F16	>0.05	0.2713
F3	0.0182	0.026	F17	>0.05	0.2897
F4	0.0425	0.204	F18	0.0201	0.2357
F5	0.0604	0.171	F19	>0.05	0.2717
F6	>0.05	0.0309	F20	>0.05	0.2990
F7	0.0051	0.2057	F21	0.0320	0.3608
F8	>0.05	0.079	F22	0.0646	0.3295
F9	0.0001	0.080	F23	>0.05	0.3975
F10	0.0001	0.034	F24	>0.05	0.4713
F11	>0.05	0.205	F25	>0.05	0.1001
F12	0.0001	0.169	F26	0.0001	0.1037
F13	0.0033	0.2287	F27	0.0001	0.2644
F14	0.0037	0.1800			

Based on the statistically significant features shown in Table 6, we selected the feature set consisting of features F1, F3, F4, F5, F7, F9, F10, F12, F13, F14, F15, F18, F21-F24, F26, F27 as the input feature vectors for the fusion-based classification. Note that all these features were selected based on the $SASMLR$ test of statistical significance except for F23 and F24, which were selected since they have relatively high information gain ratio (AIGR) among the 27 features (the second place and third place in Table 5). Our experiment compared classification accuracies using this reduced set of features to the results using the entire 27 features set, and also compared to the classification accuracies obtained in the previous research [5].

The reduced feature classifications were done for the fusion based method only to remain consistent with the previous research [5]. The classification algorithms (SVM/LDA) were applied to the reduced features, followed by fusing of the vertical segment classifications to obtain the CIN grade of the epithelium. The resulting classifications obtained in this approach are shown as confusion matrices in Table 4 for both experts.

Table 6. CIN discrimination rates for fusion-based classification using all features, whole image classification, and reduced feature set fusion-based classification for both experts

	Fusion-based Classification	
	SVM(RZ/SF)	LDA(RZ/SF)
Exact Class Label	86.9%/82.0%	88.5%/82.0%
Normal versus CIN	96.7%/90.2%	96.7%/90.2%
Off-by-One	100%/100%	98.4%/96.7%
	Whole image classification	
	SVM(RZ/SF)	LDA(RZ/SF)
Exact Class Label	67.2%/54.1%	50.8%/54.1%
Normal versus CIN	98.4%/88.6%	80.3%/88.6%
Off-by-One	90.2%/96.7%	93.4%/95.1%
	Reduced feature set	
	SVM(RZ/SF)	LDA(RZ/SF)
Exact Class Label	83.6%/83.6%	88.5%/85.3%
Normal versus CIN	96.7%/90.2%	95.1%/90.2%
Off-by-One	98.4%/100%	100%/98.4%

From Table 4, the following correct classification rates were obtained for the reduced features using the SVM-based classifier: Exact Class Label classification of 83.6%/83.6% (RZ/SF), Normal versus CIN classification of 96.7%/90.2% (RZ/SF), and Off-by-One classification of 98.4%/100% (RZ/SF). From Table 4, the correct classification rates using the LDA classifier were obtained as: Exact Class Label classification of 88.5%/85.3% (RZ/SF), Normal versus CIN classification of 95.1%/90.2% (RZ/SF), and Off-by-One classification of 100%/98.4% (RZ/SF). The highest correct classification rates obtained in previous work using the same experimental dataset and leave-one-out training/testing approach with the LDA classifier are summarized as follows [5]: Exact Class label of 70.5%, Normal versus CIN of 90.2%, and Off-by-One classification of 100%.

4.2. Analysis of Results. In this section, we use the classification results from Section IV, to compare 1) the performance among the scoring approaches, 2) the performance between the SVM and LDA classifiers, and 3) the performance between the previous research [5] and this study. Table V gives an overview of the correct recognition rates in different classification schemes examined in this paper.

From Table 6, the fusion-based classification approach shows improvement (except for same results for Normal versus CIN) compared to the whole image classification approach, when all the feature vectors are used as input for the classifiers. For the fusion-based versus whole image classification, the fusion based approach shows an improvement of 19.7% (minimum improvement from the two experts from 67.2% to 86.9%) for SVM and 27.9% (minimum improvement from the two experts from 54.1% to 82.0%) for LDA using the Exact Class Label scoring scheme. For the Normal versus CIN scoring scheme, an accuracy improvement of 1.7% (minimum improvement from the two experts from 88.5% to 90.2%) for the LDA classifier, although we note an accuracy decline of 1.7% (from 98.4% to 96.7%) was observed for the SVM classifier. For Off-by-One scoring scheme, classification accuracy increases 3.3% (minimum of two experts from 96.7% to 100%) and 1.6% (minimum of two experts from 95.1% to 96.7%) for SVM and LDA, respectively.

With feature reduction added to fusion-based classification, the fusion-based method improves in half the comparisons. Specifically, the Exact Class Label accuracy for SVM declines by 3.3% (minimum improvement from two experts of 86.9% to 83.6%) and LDA's accuracy yields zero improvement (minimum from two experts of 88.5% to 88.5%). For Normal versus CIN, there is no improvement (0%) for SVM for both experts, and a 1.6% loss (minimum from two experts of 96.7%-95.1%) in accuracy for LDA. For Off-by-One, the SVM classifier has a 1.6% decline (minimum of experts from 100% to 98.4%), and LDA has zero improvement (from 98.36% to 98.36%), and LDA has a gain of 1.6% (minimum of experts from 98.4% to 100%).

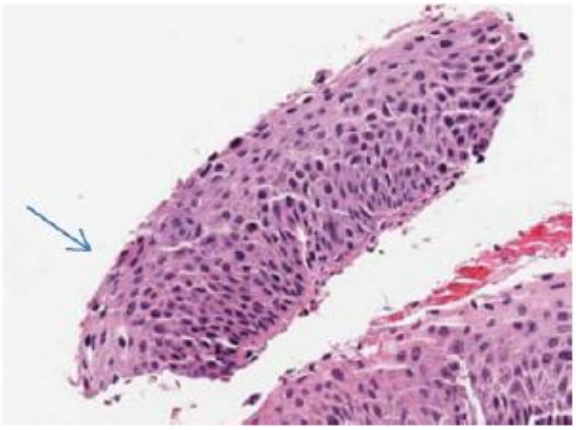

Figure 20. Misclassification example of a CIN2 image labeled as a CIN3.

Among all of classification results obtained by the two different classifiers, the highest come from the fusion-based classification. The highest Exact Class Label classification accuracy by the two experts was 88.5%/85.3% (LDA, reduced feature set). In comparison, SVM obtained 83.6%/83.6% by both experts for the reduced feature set. The accuracies for Normal versus CIN and Off-by-One are relatively high for both experts (above 90% for both SVM and LDA classifiers, and for both the full and the reduced feature sets). A summary of the results from this study and from the previous research [5] is shown in Table 5 below. Note that only the LDA classifier was reported in [5].

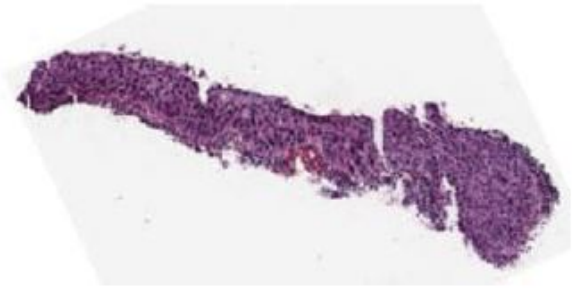

Figure 21. Misclassification example of a CIN3 image labeled as a CIN1.

In examining the classification results, the majority of the Exact Class Label classification errors are off-by-one CIN grade. This is supported with the high Off-by-One classification rates for the different experiments performed. Figure. 20 shows an example

Table 7. Summary of classification accuracies: previous research versus reduced feature set results in this paper

		Previous work	Current Work(RZ/SF)	
		LDA	SVM	LDA
Fusion-based classification	Exact Class Label	62.3%	86.9%/82.0%	88.5%/82.0%
	Normal versus CIN	88.5%	96.7%/90.2%	96.7%/90.2%
	Off-by-One	96.7%	100%/100%	98.4%/96.7%
Whole image classification	Exact Class Label	39.3%	67.2%/54.1%	50.8%/54.1%
	Normal versus CIN	78.7%	98.4%/88.5%	80.3%/88.5%
	Off-by-One	77.0%	90.2%/96.7%	93.4%/95.1%
Reuduced feature classification	Exact Class Label	70.5%	83.6%/83.6%	88.5%/85.3%
	Normal versus CIN	90.2%	96.7%/90.2%	95.1%/90.2%
	Off-by-One	100%	98.4%/100%	100%/98.4%

of an image with an expert label of CIN2 (RZ) that was labeled as a CIN3 by the LDA classifier. Inspecting Figure 20, nuclei bridge across the epithelium and are relatively uniform in density in the lower left-hand portion of the epithelium (see arrow). The nuclei features and the layer by layer DT features, particularly in the vertical segments containing the lower left-hand portion of the epithelium, provide for a higher CIN grade. In other regions of the epithelium, the nuclei density is not as uniform across the epithelium, which could provide for a less severe CIN grade label for the epithelium. Figure 21 shows an example of an image with an expert label of CIN3 (RZ) that was labeled as a CIN1 by the LDA classifier. This image has texture, nuclei distribution, and color typical of a CIN3 grade. However, the white gaps present along the epithelium were detected as acellular regions, leading to the misclassification.

The overall algorithm was found to be robust in successful identification of nuclei. Nuclei in the two lightest-stained slides and the two darkest-stained slides were counted. An average of 89.2% of nuclei in all four slides was detected. The 89.2% nuclei detection rate observed represents an advance over the results of Veta et al. [28], who found 85.5%

to 87.5% of nuclei (not strictly comparable, as these results were for breast cancer). The finding of a high percentage of nuclei in the lightest- and darkest-stained slides shows that the algorithm is adaptable and robust with regard to varying staining.

The approach in this study expands the techniques of other researchers who often process but a single cell component: the nucleus. We show in this paper that the transition from benign to cancer affects the whole cell. We have shown that not only nuclei, but in fact features of the entire cell, including inter cellular spaces, are changed due to the more rapidly growing cells. Thus, one of the top four features by p-value is the proportion of regions of cytoplasm in the image (F12).

We also sought to use layers to better represent the CIN transition stages. The number of DTs in the middle layer was also one of the top four features by p-value, validating our approach of analysis by layers. The last two features with the most significant p-values were the energy of a GLCM (the sum of squared elements in GLCM in horizontal and vertical directions). The energy in the GLCM appears to capture the growing biological disorder as the CIN grade increases.

We emphasize that between the previous research [9] and our paper, 1) the training and testing datasets are the same; 2) the classifier (LDA) is the same and we investigated the SVM classifier; and 3) the scoring schemes, (Exact Class Label, Normal versus CIN, and Off-by-One) are the same (in the previous research, Off-by-One was called "windowed class"). There are two differences between the previous and current work. First, CIN classification results are reported for two experts (RZ and SF) in this study to demonstrate CIN classification improvement over previous work, even with variations in the expert CIN truthing of the experimental dataset. Second, three acellular features (F18, F21, F22) and two layer-by-layer triangle (F26, F27) nuclei features were found to be significant (from Table 5), which are new in this paper, contribute to improved CIN discrimination capability over previous work.

For the fusion-based method applied to all the feature vectors, Table 5 shows that Exact Class Label accuracy increases by 19.7% (from 62.3% to 82.0% for the lower of the two expert results) for LDA. For the whole image method, LDA improved by 14.8% (from 39.3% to 54.1% for the lower of the two expert results). For fusion-based classification with reduced feature vectors, accuracy increases by 14.8% (from 70.5% to 85.3% for the lower of the two experts) for LDA. Since Exact Class Label is the most stringent of the scoring schemes we used, we interpret these results as showing a substantial gain in classification in classification accuracy when using the nuclei features and nuclei-related features.

The Off-by-One classification achieved excellent classification accuracy (100%) with the SVM and LDA classifiers, which matches the results from the previous study [5]. This classification metric gives more evidence of the similarity of neighboring classes (Normal/CIN1 or CIN1/CIN2 or CIN2/CIN3) and the difficulties in discriminating between them [21, 12, 22]. It is also consistent with the intra- and interpathologist variation in labeling of these images. The two experts for this study differed in the CIN labeling of five images (out of 61) (or 8.2%) in the experimental dataset, with the experts differing by only 1 CIN grade (higher or lower) in each of the five cases.

Overall, the 88.5%/85.3% accuracy by the two experts of Exact Class Label prediction using the reduced features is 23.0% higher than published results for automated CIN diagnosis (62.3%) as presented by Keenan et al. in [14], 17.3% higher than the accuracy of the method used by Guillaud et al. (68%) in [8], and 14.8% higher than the accuracy of the method by De [5], although we note that only in the comparison with De were the same training and testing sets used. The classification results presented in this study for the two experts only differed by greater than 8.2% in the Exact Class Label of the whole image. Thus, the experimental results suggest that the involvement of nuclei and nuclei-related features using vertical segment classification and fusion for obtaining the image-based CIN

classification is an improvement over the existing methods for automated CIN diagnosis. Even though our method outperformed published results, we note that there is potential for further improvement.

5. SUMMARY AND FUTURE WORK

In this study, we developed new features for the automated CIN grade classification of segmented epithelium regions. New features include nuclei ratio features, acellular area features, combination features, and layer-by-layer triangle features. We carried out epithelium image classification based on these ground truth sets: 1) two experts labeled 62 whole epithelium images as Normal, CIN1, CIN2, and CIN3, and 2) investigator labeling of ten vertical segments within each epithelium image into the same four CIN grades. The vertical segments were classified using an SVM or LDA classifier, based on the investigator-labeled training data of the segments with a leave-one-out approach. We used a novel fusion-based epithelium image classification method which incorporates a voting scheme to fuse the vertical segment classifications into a classification of the whole epithelium image. We evaluated the classification results with three scoring schemes, and compared the classification differences by classifiers, by scoring schemes, and the classification results of this paper as compared to our previous work [5].

We found that the classification accuracies yielded in this study with nuclei features outperformed that of the previous work [5]. Using the LDA classifier upon the reduced set of features, and based on an Off-by-One classification scoring scheme for epithelium region classification, correct prediction rates as high as 100% were obtained. Normal versus CIN classification rates were as high as 96.72%, whereas the rates for Exact Class Labels were as high as 88.52% using a reduced set of features. Future research may include the use of adaptive critic design methods for classification of CIN grade. Also, it is important to

include more cervix histology images to obtain a comprehensive dataset for different CIN grades. With the enhancement of the dataset, inter or intra pathologist variations can be incorporated [21].

Gwilym Lodwick, among his many contributions to diagnostic radiology, contributed to our basic knowledge of pattern recognition by both humans and computers. The importance of diagnostic signs, which he also termed mini patterns, was stated: "Signs, the smallest objects in the picture patterns of disease, are of vital importance to the diagnostic process in that they carry the intelligence content or message of the image" [18]. In this context, Professor.Lodwick also maintained that these signs are at the heart of the human diagnostic process. The results of our study appear to indicate that the new layer by layer and vertical segment nuclei features, in the domain of cervical cancer histopathology, provide useful signs or mini patterns to facilitate improved diagnostic accuracy. With the advent of advanced image processing techniques, these useful signs may now be employed to increase the accuracy of computer diagnosis of cervical neoplasia, potentially enabling earlier diagnosis for a cancer that continues to exact a significant toll on women worldwide.

6. ACKNOWLEDGMENTS

The authors would like to thank the medical expertise and collaboration of Dr. M. Schiffman and Dr. N.Wentzensen, both from the National Cancer Institute's Division of Cancer Epidemiology and Genetics.

REFERENCES

[1] Agresti, A., *An introduction to categorical Data Analysis*, volume 22, 1996, ISBN 9780471226185, doi:10.1198/jasa.2008.s251.

[2] Banerjee, K., 'Uterine cervical cancer histology image feature extraction and classification,' 2014.

[3] Borovicka, J., 'Circle detection using Hough transforms Course Project: COMS30121- Image Processing and Computer Vision,' Technical report, 2003.

[4] Chang, C.-C. and Lin, C.-j., 'LIBSVM: a library for support vector machines,' ACM Transactions on Intelligent Systems and Technology (TIST), 2011, **2**(3), pp. 1–39, ISSN 21576904, doi:10.1145/1961189.1961199.

[5] De, S., Stanley, R. J., Lu, C., Long, R., Antani, S., Thoma, G., and Zuna, R., 'A fusion-based approach for uterine cervical cancer histology image classification,' Computerized Medical Imaging and Graphics, 2013, **37**(7-8), pp. 475–87, ISSN 1879-0771, doi:10.1016/j.compmedimag.2013.08.001.

[6] Fan, R.-E., Chen, P.-H., and Lin, C.-J., 'Working Set Selection Using Second Order Information for Training Support Vector Machines,' J. Mach. Learn. Res., 2005, **6**, pp. 1889–1918, ISSN 1532-4435, doi:http://dx.doi.org/10.1016/j.patcog.2012.03.004.

[7] Gonzalez, R. and Woods, R., *Digital image processing*, 2002, ISBN 0201180758, doi:10.1016/0734-189X(90)90171-Q.

[8] Guillaud, M., Adler-Storthz, K., Malpica, A., Staerkel, G., Matisic, J., Van Niekirk, D., Cox, D., Poulin, N., Follen, M., and MacAulay, C., 'Subvisual chromatin changes in cervical epithelium measured by texture image analysis and correlated with HPV,' Gynecologic Oncology, 2005, **99**(3 SUPPL.), pp. 16–23, ISSN 0090-8258, doi: 10.1016/j.ygyno.2005.07.037.

[9] Hall, M. A., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H., 'The WEKA data mining software: an update,' SIGKDD Explorations, 2009, **11**(1), pp. 10–18, ISSN 19310145, doi:10.1145/1656274.1656278.

[10] He, L., Long, L. R., Antani, S., and Thoma, G., 'Computer assisted diagnosis in histopathology,' Sequence and Genome Analysis: Methods and Applications, 2010, **3**, pp. 271–287.

[11] Hosmer, D. W. and Lemeshow, S., *Applied Logistic Regression*, 1, 2000, ISBN 9780471722144, doi:10.2307/2074954.

[12] Ismail, S. M., Colclough, A. B., Dinnen, J. S., Eakins, D., Evans, D. M., Gradwell, E., O'Sullivan, J. P., Summerell, J. M., and Newcombe, R., 'Reporting cervical intraepithelial neoplasia (CIN): intra- and interpathologist variation and factors associated with disagreement.' Histopathology, 1990, **16**(4), pp. 371–376, ISSN 0309-0167 (Print).

[13] Jeronimo, J., Schiffman, M., Long, R. L., Neve, L., and Antani, S., 'A tool for collection of region based data from uterine cervix images for correlation of visual and clinical variables related to cervical neoplasia,' in 'Proceedings. 17th IEEE Symposium on Computer-Based Medical Systems,' IEEE Comput. Soc, ISBN 0-7695-2104-5, ISSN 1063-7125, 2004 pp. 558–562, doi:10.1109/CBMS.2004.10.

[14] Keenan, S. J., Diamond, J., Glenn McCluggage, W., Bharucha, H., Thompson, D., Bartels, P. H., and Hamilton, P. W., 'An automated machine vision system for the histological grading of cervical intraepithelial neoplasia (CIN),' Journal of Pathology, 2000, **192**(3), pp. 351–62, ISSN 0022-3417, doi:10.1002/1096-9896(2000)9999:9999<::AID-PATH708>3.0.CO;2-I.

[15] Krzanowski, W. J., *Principles of Multivariate Analysis: A UserâĂŹs Perspective*, Oxford University Press, New York, Oxford, 1988.

[16] Kumar, V., Abbas, A. K., Fausto, N., and Aster, J. C., *Robbins and Cotran Pathologic Basis of Disease*, Elsevier Health Sciences, 2014.

[17] Li, T., Zhu, S., and Ogihara, M., 'Using discriminant analysis for multi-class classification: An experimental investigation,' Knowledge and Information Systems, 2006, **10**(4), pp. 453–472, ISSN 02191377, doi:10.1007/s10115-006-0013-y.

[18] Lodwick, G. S., 'Diagnostic Signs and Minipatterns,' Proc Annu. Symp. Comput. Appl. Ded.Care, 1980, **3**, pp. 1849–1850, ISSN 0195-4210 (Print).

[19] Mathworks, 'Matlab (2015),' ????

[20] Maurer, C. and Raghavan, V., 'A linear time algorithm for computing exact Euclidean distance transforms of binary images in arbitrary dimensions,' IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, **25**(2), pp. 265–270, ISSN 0162-8828, doi:10.1109/TPAMI.2003.1177156.

[21] McCluggage, W. G., Walsh, M. Y., Thornton, C. M., Hamilton, P. W., Date, A., Caughley, L. M., and Bharucha, H., 'Inter- and intra-observer variation in the histopathological reporting of cervical squamous intraepithelial lesions using a modified Bethesda grading system,' British journal of obstetrics and gynaecology, 1998, **105**(2), pp. 206–10, ISSN 0306-5456.

[22] Molloy, C., Dunton, C., Edmonds, P., Cunnane, M. F., and Jenkins, T., 'Evaluation of colposcopically directed cervical biopsies yielding a histologic diagnosis of CIN 1,2.' Journal of lower genital tract disease, 2002, **6**(2), pp. 80–3, ISSN 1089-2591.

[23] Pal, M., 'Multinomial logistic regression-based feature selection for hyperspectral data,' International Journal of Applied Earth Observation and Geoinformation, 2012, **14**(1), pp. 214–220, ISSN 15698432, doi:10.1016/j.jag.2011.09.014.

[24] Preparata, F. P. and Shamos, M. I., *Computational geometry: an introduction*, volume 47, 1985, ISBN 0387961313, doi:10.2307/2008200.

[25] Rao, C., Toutenburg, R., Heumann, H., and Christian, S., *Linear Models and Generalizations Least Squares and Alternatives*, 2008, ISBN 9783540742265, doi: 10.1007/s10182-008-0101-2.

[26] Rudas, I. J., 'Detection of abrupt changes. Theory and application,' Control Engineering Practice, 1994, **2**(4), pp. 729–730, ISSN 09670661, doi:10.1016/0967-0661(94)90196-1.

[27] Van Der Marel, J., Quint, W. G. V., Schiffman, M., van de Sandt, M. M., Zuna, R. E., Dunn, S. T., Smith, K., Mathews, C. A., Gold, M. A., Walker, J., and Wentzensen, N., 'Molecular mapping of high-grade cervical intraepithelial neoplasia shows etiological dominance of HPV16,' International Journal of Cancer, 2012, **131**(6), pp. 946–53, ISSN 00207136, doi:10.1002/ijc.27532.

[28] Veta, M., van Diest, P. J., Kornegoor, R., Huisman, A., Viergever, M. A., and Pluim, J. P. W., 'Automatic Nuclei Segmentation in H&E Stained Breast Cancer Histopathology Images,' PloS one, 2013, **8**(7), p. e70221, ISSN 19326203, doi: 10.1371/journal.pone.0070221.

[29] Wang, Y., Crookes, D., Eldin, O. S., Wang, S., Hamilton, P., and Diamond, J., 'Assisted diagnosis of cervical intraepithelial neoplasia (CIN),' IEEE Journal on Selected Topics in Signal Processing, 2009, **3**(1), pp. 112–121, ISSN 1932-4553, doi: 10.1109/JSTSP.2008.2011157.

[30] World Health Organization, 'Comprehensive Cervical Cancer Control. A guide to essential practice,' WHO Library Cataloguing-in-Publication Data, 2006, pp. 1–284.

II. ENHANCEMENTS IN LOCALIZED CLASSIFICATION FOR UTERINE CERVICAL CANCER DIGITAL HISTOLOGY IMAGE ASSESSMENT

P. Guo[1] H. Almubarak[1] R.Joe Stanley[1] Rodney Long[2] Sameer Antani[2] George Thoma[2]

Rosemary Zuna[3] Shelliane R. Frazier[4] Randy H. Moss[1] William V.Stoecker[5]

[1]Department of Electrical & Computer Engineering

Missouri University of Science and Technology, Rolla

[2]Lister Hill National Center for Biomedical Communications,

National Library of Medicine, National Institutes of Health, DHHS, Bethesda, MD

[3]Department of Pathology,

University of Oklahoma Health Sciences Center, Oklahoma City, OK

[4]Surgical Pathology Department,

University of Missouri Hospitals and Clinics, Columbia, MO

[4]Stoecker & Associates, Rolla, MO, USA

Email: stanleyj@mst.edu

ABSTRACT

Background: In previous research, we introduced an automated, localized, fusion-based approach for classifying uterine cervix squamous epithelium into Normal, CIN1, CIN2, and CIN3 grades of cervical intraepithelial neoplasia (CIN) based on digitized histology image analysis. As part of the CIN assessment process, acellular and atypical cell concentration features were computed from vertical segment partitions of the epithelium region to quantize the relative distribution of nuclei. Methods: Feature data was extracted from 610 individual segments from 61 images for epithelium classification into categories of Normal, CIN1, CIN2, and CIN3. The classification results were compared against CIN labels obtained from two pathologists who visually assessed abnormality in the digitized

histology images. In this study, individual vertical segment CIN classification accuracy improvement is reported using the logistic regression classifier for an expanded data set of 118 histology images. Results: We analyzed the effects on classification using the same pathologist labels for training and testing versus using one pathologist's labels for training and the other for testing. Based on a leave-one-out approach for classifier training and testing, exact grade CIN accuracies of 81.29% and 88.98% were achieved for individual vertical segment and epithelium whole image classification, respectively. Conclusions: The Logistic and Random Tree classifiers outperformed the benchmark SVM and LDA classifiers from previous research. The Logistic Regression classifier yielded an improvement of 10.17% in CIN Exact grade classification results based on CIN labels for training-testing for the individual vertical segments and the whole image from the same single expert over the baseline approach using the reduced features. Overall, the CIN classification rates tended to be higher using the training-testing labels for the same expert than for training labels from one expert and testing labels from the other expert. The Exact class fusion-based CIN discrimination results obtained in this study are similar to the Exact class expert agreement rate.

Keywords: Cervical cancer, cervical intraepithelial neoplasia, fusion-based classification, image processing.

1. INTRODUCTION

There were 528,000 new invasive cervical cancer cases and an estimated 266,000 deaths reported worldwide in 2012[1]. Screening tests to detect cervical cancer and its precursor lesions include Pap, colposcopy to visually inspect the cervix, and microscopic interpretation of histology slides by a pathologist when biopsied cervix tissue is available. Microscopic evaluation of histology slides by a qualified pathologist has been used as a standard of diagnosis. The pathologist visually inspects the slide for the presence of cervical intraepithelial neoplasia (CIN), a premalignant condition in the epithelium. Figure

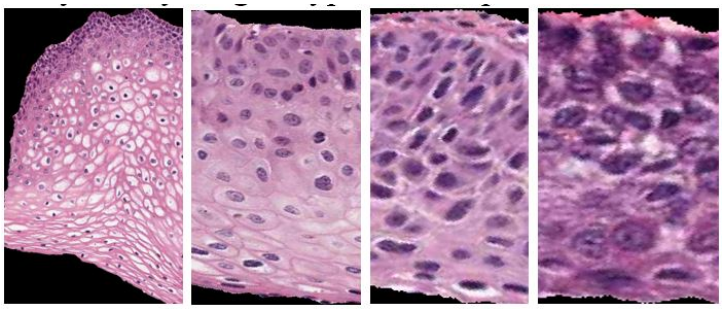

Figure 1. Cervical intraepithelial neoplasia grade label examples highlighting the increase of immature atypical cells from epithelium bottom to top with increasing cervical intraepithelial neoplasia severity. (a) Normal, (b) cervical intraepithelial neoplasia 1, (c) cervical intraepithelial neoplasia 2, (d) cervical intraepithelial neoplasia 3 .

1 shows examples of the CIN grades normal, CIN1, CIN2, and CIN3. CIN1 corresponds to mild dysplasia (abnormal change), whereas CIN2 and CIN3 are used to denote moderate dysplasia and severe dysplasia, respectively. Histologic criteria for CIN include increasing immaturity and cytologic atypia in the epithelium.

As CIN increases in severity, the epithelium has been observed to show delayed maturation with an increase in immature atypical cells from bottom to the top of the epithelium [1]. As shown in Figure 1, atypical immature cells are seen mostly in the bottom third of the epithelium for CIN1 [Figure 1b]. For CIN2, the atypical immature cells typically appear in the bottom two-thirds of the epithelium [Figure 1c], and for CIN3, atypical immature cells typically are found in the full thickness of the epithelium [Figure 1d]. In addition to analyzing the progressively increasing quantity of atypical cells from bottom to top of the epithelium, identification of nuclear atypia is also significant [1]. Nuclear atypia are characterized by nuclei of abnormal shapes and sizes within the epithelium region. Visual assessment of this nuclear atypia may be difficult, due to a large number of nuclei present and tissue heterogeneity. This may contribute to diagnostic inter- and intra-pathologist variation.

Computer-assisted methods (digital pathology) have been explored for CIN diagnosis in other studies and provided the foundation for the work reported [2]. In depth literature reviews for related studies have been presented [3, 2]. In addition, this paper builds off techniques for semi-automated CIN assessment for epithelium regions in digitized pathology images examining texture features, nuclei determination and Delaunay triangulation analysis[4, 5], medial axis determination, and localized CIN grade assessment. This paper extends the study [3, 2], for the development of image analysis and classification techniques for individual vertical segments obtained from partitioning the epithelium along the medial axis. A logistic regression classifier is explored for CIN classification for comparison with support vector machine (SVM) and linear discriminant analysis (LDA) classifier approaches for individual vertical segment classification. CIN grades from two pathologists for 118 digitized histology images are used as ground truth for CIN classification accuracy.

The order of the remaining sections of the article is as follows: Section II presents the image analysis and classification approaches used in this research; Section III describes the experiments performed; Section IV presents and analyzes the results obtained and a discussion; Section V provides the study conclusions.

2. METHODS

Figure 2 presents an overview of the approach for analyzing the digitized pathology epithelium images:

Step 1. Detect the medial axis of the segmented epithelium region

Step 2. Divide the segmented image into 10 vertical segments orthogonal to the medial axis

Step 3. Extract features from each of the vertical segments

Step 4. Use the classification algorithms to classify each segment into one of the CIN grades

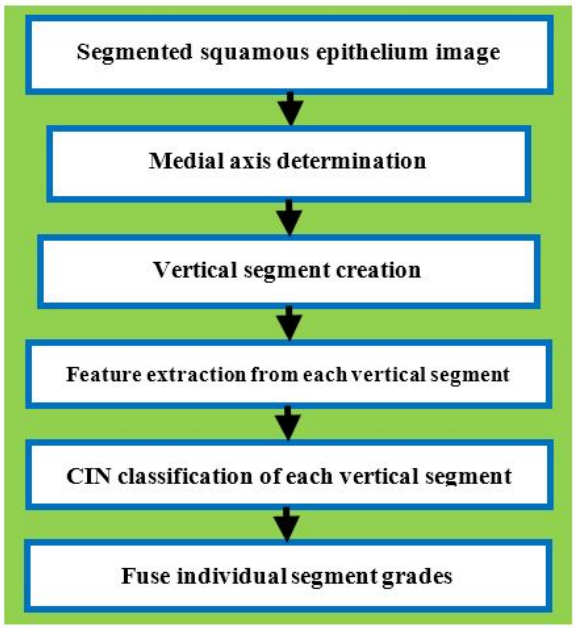

Figure 2. Digitized pathology epithelium image analysis procedures.

Step 5. Fuse the CIN grades from every ten vertical segments in one image to obtain the CIN grade of the whole epithelium.

This approach was used in previous studies [3, 2]. The following sections present each step in detail.

2.1. Pathologist Ground Truth Description. For the image and CIN classification techniques explored in this research, 118 full-color digitized histology images are used with H and E stain preparations of tissue sections of normal cervical tissue and three grades of cervical carcinoma in situ. This data set extends the 61 images used in previous studies [3, 2]. In this study, expert pathologists (RZ, SF) provided CIN grades for the whole epithelium image and for the 10 vertical segments into which each image was partitioned [Figure 3 and Table 1].

Note that, the CIN grades from the expert pathologists for the individual vertical segments within an image sometimes vary between the experts and that the CIN grades for the individual vertical segments can be different from the whole image. The ground

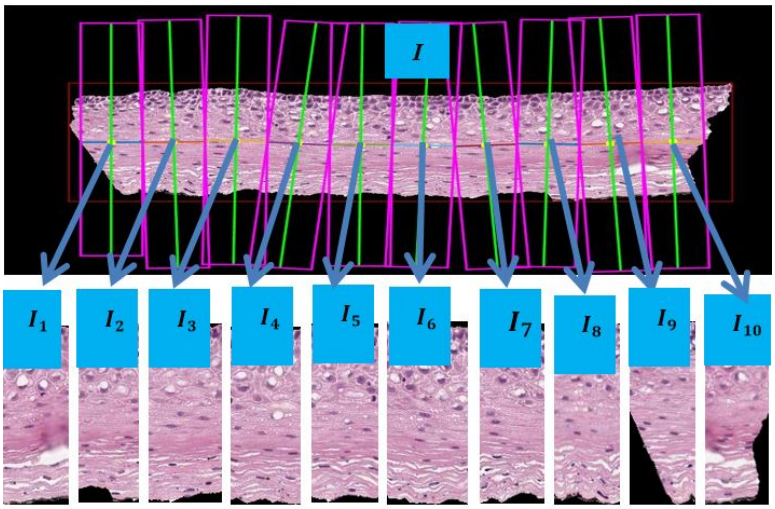

Figure 3. Epithelium image example with vertical segment images (I_1, I_2, I_2,...,I_{10}) determined from bounding boxes after dividing the medial axis into ten line segment approximations after medial axis computation.

truth is given as two groups of CIN grades for every segment out of our 118-image data set. Each single label is specified as a class number to denote the dysplasia and severe dysplasia. In this case, the pathologists gave "1" as Normal, "2" as CIN1, "3" as CIN2 and "4", as CIN3. A pathologist labeled a vertical segment "0" if the pathologist was not able to make any CIN grade decision due to insufficient image information or detail (the 9th segment in image 2 [RZ] and the 9th and 10th segment in image 4 [SF] [Table 1]). Since 118 digitized histology images are used in this study to create vertical segments for feature extraction and classification, 1180 segments in total are labeled by both pathologists to generate two groups of ground truth, respectively. Table 1 provides CIN labels from both pathologists (RZ/SF) for the 10 vertical segments from 10 histology images as examples of the experimental data set.

Table 1 shows that the two pathologists agree with each other on some of the segments and disagree on others. For example, from image 8, RZ assigns every segment as CIN3 (4), but SF only labels the 3rd, 4th, 6th, and the 9th as CIN3 (4) with the others as CIN2 (3). Part of the rationale for this paper is to show that the classification results for

Table 1. Ground truth cervical intraepithelial neoplasia grade labels for both experts

Image name	Individual segment classifications (RZ/SF)										Image classification(RZ/SF)
	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	6^{th}	7^{th}	8^{th}	9^{th}	10^{th}	
1	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2
2	4/4	3/4	3/3	4/3	3/3	4/3	3/4	3/4	0/4	4/3	3/3
3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3
4	4/4	4/4	4/4	4/3	4/2	4/2	4/2	4/3	4/0	4/0	4/4
5	3/2	3/2	3/2	3/2	3/2	3/2	3/2	3/3	3/3	3/3	3/3
6	1/1	1/1	1/1	1/1	1/1	1/1	1/1	1/1	1/1	1/1	1/1
7	4/3	4/3	4/3	4/3	4/3	4/3	4/3	4/3	4/4	4/4	4/4
8	4/3	4/3	4/4	4/4	4/3	4/4	4/3	4/3	4/4	4/3	4/4
9	3/2	3/3	3/4	3/3	3/3	3/4	3/3	3/4	3/3	3/4	3/3
10	3/1	3/1	3/2	3/1	2/2	2/1	2/2	2/2	3/3	3/2	3/3

the individual vertical segments and the whole image are within the variation of the expert pathologist designations and that there is inter-pathologist variation within an image and for the image-based classification[Table 1].

Three methods were used for assigning "truth labels" to the individual vertical segments, including the "0" labeled segments that the pathologists did not label, producing three (slightly) different sets of ground truth labels for evaluating the classification algorithms developed. The three methods examined to determine ground truth labels are:

1. Use the image label for every single segmentation regardless of the individual labels, which are denoted as "Image Label"

2. Keep the pathologist labels for the non-"0" segments and replace the "0" segments with the majority of individual labels by the pathologist within these 10 segments, which are denoted as "Major Sub"

3. Keep the pathologist labels for the non-"0" segments and replace the "0" segments with the whole image label by the pathologist, which are denoted as "Image Sub."

2.2. Medial Axis Detection and Segments Creation. The method for computing the medial axis, which is based on the distance transform, is presented in detail[2].The resulting medial axis is partitioned into ten segments of approximately equal length, perpendicular line slopes are estimated at the mid-points of each segment, and vertical lines are projected at the end points of each segment to generate ten vertical segments for analysis. The epithelium image is partitioned into ten vertical segments to facilitate localized diagnostic classification on sub-regions within the epithelium.

2.3. Feature Extraction. Features are computed for each of the ten vertical segments of the whole image, I_1, I_2, I_2,...,I_{10}. All the segments of one whole image are feature-extracted in a sequence, from left to right, I_1- I_{10} [Figure 3]. These features were developed in previous research[2]. A summary of those features is presented here. In total, five different types of features were computed, including: (1) Texture features (F1-F10)[3], (2) cellularity features (F11-F13), (3) nuclear features (F14, F15), (4) acellular (light area) features (F16-F22), (4) combination features (F23, F24), and (5) advanced layer-by-layer triangle features (F25-F27)[2].

2.3.1. Texture and Color Features. The texture and color features were used in our previous work and are described[2]. The texture features include contrast (F1), energy (F2), correlation (F3), and homogeneity (F4) of the segmented region, combined with the same statistics (contrast, energy, and correlation) generated from the gray level co-occurrence matrix (GLCM) of the segment (F5-F10). These features are generated using the statistics of the GLCM matrix[2, 6, 7] to describe the contrast and the uniformity of the region.

2.3.2. Nuclear Features. The dark shading color feature discussed in the previous research[2] corresponds to nuclei, which appear within epithelial cells in various shapes and sizes. Nuclei tend to increase in both number and size as the CIN level increases[1]. This linkage between nuclear characteristics and CIN levels motivates our development of

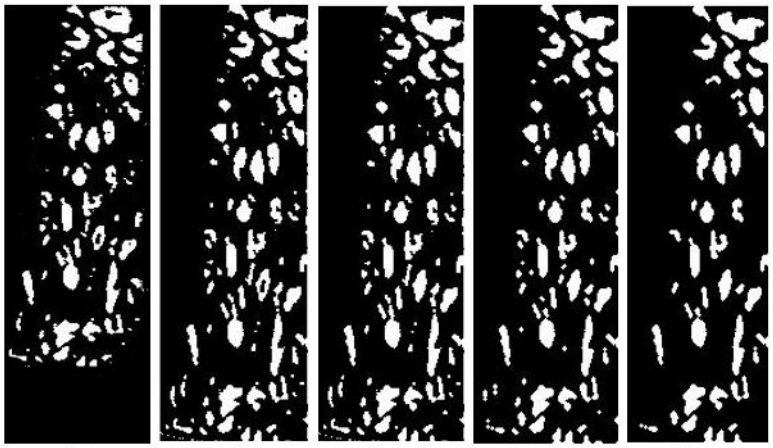

Figure 4. Image examples of nuclei detection algorithm. (a) Image with preliminary nuclei objects obtained from clustering (Step 1). (b) Image closing to connect nuclei objects (Step 2). (c) Image with hole filling to produce nuclei objects (Step 3). (d) Image opening to separate nuclei objects (Step 4). (e) Image with nonnuclei (small) objects eliminated (Step 5) .

algorithms for nuclei detection feature extraction. In this research, the algorithms of nuclei detection and nuclear feature extraction are developed to obtain features to facilitate CIN classification. Specifically, the following steps are performed [Figure 5]:

Step 1. Cluster the histogram-equalized image into clusters of background (darkest), nuclei and lighter (lightest) epithelium regions using the K-means algorithm (K = 4). Generate a mask image containing the pixels closest to the nuclei cluster (second darkest)

Step 2. Use the Matlab function *imclose* with a circular structuring element of radius 4 to perform morphological closing on the nuclei mask image

Step 3. Fill the holes in the image from Step 2 with Matlab's *imfill* function for this process

Step 4. Use the Matlab's imopen to perform morphological opening with a circular structuring element of radius 4 on the image from Step 3

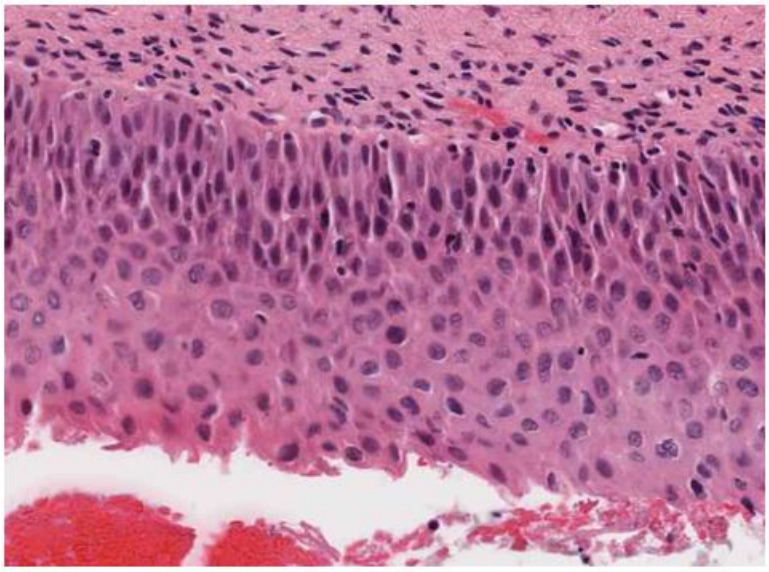

Figure 5. Misclassification example of a cervical intraepithelial neoplasia 2 image labeled as a cervical intraepithelial neoplasia 3.

Step 5. Eliminate small area noise objects (nonnuclei objects) within the epithelium region of interest from the mask in Step 4, with the area opening operation using the Matlab function *bwareaopen*.

2.3.3. Acellular Features. Extracting the light area regions is challenging due to the color and intensity variations in the epithelium images. Each of the L*, a*, and b* planes of CIELAB color space were evaluated for characterizing the light areas. It was empirically determined that L* provides the best visual results. The following outlines the methods used to segment the histology images:

Step 1. Convert the original image from RGB color space to L* a* b* color space, then select the luminance component L* [Figure 4a]

Step 2. Perform adaptive histogram equalization on the image from Step 1 using Matlab's $adapthisteq$. $Adapthisteq$ operates on small regions (tiles)[2] for contrast enhancement so that the histogram of the output region matches a specified histogram and combines neighboring tiles using bilinear interpolation to eliminate artificially induced boundaries [Figure 4b]

Step 3. After the image has been contrast-adjusted, the image is binarized by applying an empirically determined threshold of 0.6. This step is intended to eliminate the dark nuclear regions and to retain the lighter nuclei and epithelium along with the light areas [Figure 4c]

Step 4. Segment the light areas using the K-means algorithm based on [3, 8], with K = 4. The K-means algorithm input is the histogram-equalized image from Step 2 multiplied by the binary thresholded image from Step 3. A light area clustering example is given in Figure 4d.

Step 5. Remove from the image all objects having an area <100 pixels, determined empirically, using the Matlab function $regionprops$[9]. A morphological closing is performed with a disk structure element of radius 2. An example result is shown in Figure 4e.

2.3.4. Combination Features. After both the nuclear features and the acellular features were extracted, combination features were calculated with the intent to capture the relative increase in nuclei numbers as CIN grade increases. One is the ratio of the acellular number to the nuclei number (F23), and the other is the ratio of the acellular area to the total nuclei area (F24).

2.3.5. Triangle Features. In this research, the Delaunay triangle method was used, but restrict the geometrical regions it can act upon, as follows. Before forming the Delaunay triangles,[4, 10] with the vertices provided by the nuclei detection results from nuclear feature section, the vertical segment being processed is sub-divided into three vertical layers, as illustrated in Figure 16 in [2]. The aim is to associate the presence of increasing

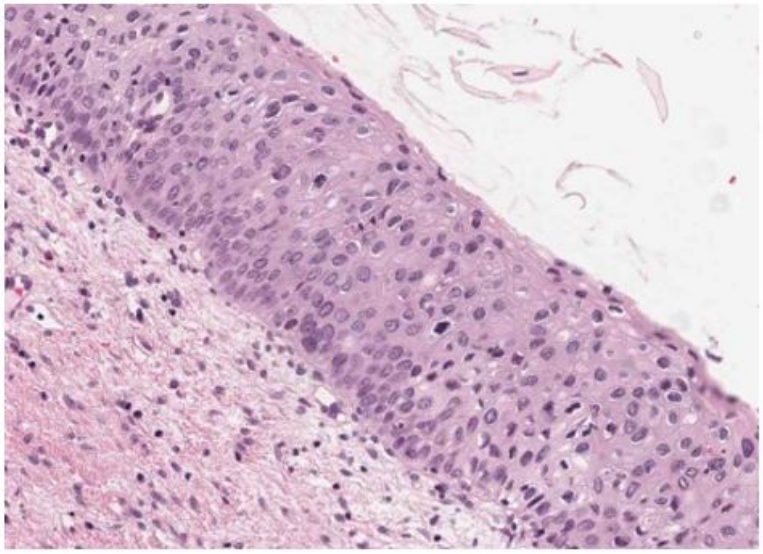

Figure 6. Misclassification example of a cervical intraepithelial neoplasia 2 image labeled as a cervical intraepithelial neoplasia 1.

nuclei throughout the epithelium with increasing CIN grades, namely: abnormality of the bottom third of the epithelium roughly corresponds to CIN1; abnormality of the bottom two-thirds, to CIN2; and abnormality of all three layers, to CIN3. These layers are referred to as the bottom, mid, and top.

3. EXPERIMENT PERFORMED

Experiments were performed using the data set consisting of 118 digitized histology images, which were CIN labeled by two experts (RZ and SF) (RZ: 38 normal, 26 CIN1, 26 CIN2, and 26 CIN3; SF: 40 normal, 25 CIN1, 24 CIN2, and 29 CIN3).

3.1. Fusion-based Cervical Intraepithelial Neoplasia Grade Classification of Vertical Segment Images Labeling.

Labeling. The experimental data set consists of 118 digitized histology images with vertical segments obtained using the medial axis detection and vertical segment partitioning algorithms presented in Section II. An additional image from previous research [2] was used

for image processing parameter determination. For this experimental data set, CIN grades were obtained for the 10 vertical segments from each image from both pathologists (RZ/SF), and the image analysis, feature extraction and classification approaches presented in Section II were performed for each vertical segment image. Epithelium image truth labels from both pathologists (SF and RZ) are used as training and testing labels, unlike our previous study[2] where only one expert's truth label was used in training and testing. As described in Section IIA, the pathologists were unable to assign labels for some segments, for which the Image Label, Major Sub, and Image Sub methods were used to generate "truth labels" for these segments (Section II for definitions).

Classification. For CIN discrimination, all the features extracted from the vertical segment images were used as inputs to SVM, LDA, logistic regression, and random forest classifiers. The LIBSVM[11] implementation of the SVM and LDA classifiers were used, as in our previous study.[4] The SVM implementation uses a linear kernel with four weights as the fractions of the images in each CIN class (normal, CIN1, CIN2, and CIN3). Logistic regression had a multinomial logistic regression (MLR) model for predicting probabilities for each class [12]. Random forest [13, 14], used combinations of tree predictors such that each tree depends on the values of a random vector sampled with the same distribution for all trees in the forest [14, 15].

Individual features were normalized by subtracting the mean training set feature value and dividing by the standard deviation training set feature value [2]. In this approach, the classifier is trained based on the individual segment feature vectors for all but the left-out epithelium image (test image), which was called "leave-one-out" approach [2]. Four approaches were explored for using the CIN expert truth labels for the individual vertical segments and the whole epithelium image for classifier algorithm training and testing, including:

1. SF-SF CIN labels as training-testing labels

2. RZ-RZ CIN labels as training-testing labels

3. RZ-SF CIN labels for training-testing sets

4. SF-RZ CIN labels for training-testing sets.

Using the different pathologist CIN label training and test combinations for vertical segment classification, the influence on inter- and intra-pathologist CIN truth labels was examined on individual vertical segment classification accuracy.

Then, the predicted CIN grades of the vertical segment images were fused to obtain the CIN grade of the entire test epithelium image. The fusion of the CIN grades of the vertical segment images was completed using a voting scheme [2]. The CIN grade of the test image was assigned to the most frequently occurring class over the ten vertical segments. If a tie was found among the most frequently appearing case of CIN grades, the test image was assigned with the higher/highest one (i.e., the most severe diagnostic grade). For example, if there was a tie between CIN2 and CIN3, then the image was assigned CIN3. As previously explained in Section II, there were four different ways of conducting our experiments by using different combinations of the labels from the two pathologists for training and testing. This resulted in four different groups of classification results. The input images for each group are the same 118 histological images.

Scoring schemes. Two scoring schemes were used for evaluating the results. Specifically, the same schemes were utilized as [2] for compatibility with those results. The schemes are summarized below: Scheme 1 (exact class label): The first approach is exact classification, which means that a label was considered correct if and only if the class label assigned to the test image by our algorithm was the same as the ground truth label. Scheme 2 (normal vs. CIN): For the second scoring scheme, the classification result was considered correct if and only if when a ground truth Normal grade was classified Normal by our algorithm and a ground truth CIN (1-3) grade was classified as CIN by our algorithm.

3.2. Feature Evaluation and Selection. In [2], a *SAS* implementation of MLR [13, 16, 17, 18, 19] and a *Weka* attribute information gain evaluator were utilized for feature selection. MLR was used for modeling nominal outcome variables, and the P values

obtained from the MLR output were utilized as criteria for selecting features when the P value is less than an appropriate alpha (α) value[16, 17, 18, 19]. For $Weka$ analysis, the features are ranked in an order by "attributes information gain ratio" where the higher the ratio, the more significant the feature will be[2]. Both feature evaluation methods are applied in this study to improve the classification outcomes as well as to keep the classification results comparable to the study by Guo et al[2]. Feature selection was done based on the whole image labels of RZ applied to the individual vertical segments.

4. EXPERIMENTAL RESULTS AND ANALYSIS

4.1. Experimental Results. As explained in the previous section, the vertical segment image classifications (CIN grading) were obtained using SVM, LDA, logistic, and random forest classifiers with a leave-one-image-out approach based on all the twenty-seven features generated for each vertical segment. This yields classification labels for each of the 10 vertical segments in an epithelium image. Then, the CIN classification for the whole epithelium image is obtained by fusing the vertical segment labels using a voting scheme. The performance of these epithelium image classifications was evaluated using the two scoring schemes presented in Section III.

Table 2. Individual vertical segment exact class label classification results based on all 27 features using same expert labels for training-testing sets (RZ-RZ and SF-SF)

	SVM (RZ-RZ/SF-SF)	LDA (RZ-RZ/SF-SF)	Logistic (RZ-RZ/SF-SF)	Random tree (RZ-RZ/SF-SF)
Imagelabel (%)	62.71/64.92	60.16/63.23	81.29/80.10	78.38/76.18
Image sub (%)	69.32/70.59	71.10/69.52	75.64/76.27	76.27/75.42
Major sub (%)	69.40/69.58	69.52/71.61	74.23/73.64	73.39/71.52

For the first set of experiments, individual vertical segment classification is examined. Each individual vertical segment is assigned a CIN grade label using the Image Label, Major Sub, and Image Sub approaches from Section IIA for training and testing the SVM, LDA, logistic regression, and random forest classifiers. For these experiments, individual

Table 3. Features with corresponding p-values and attribute information gain ratio

Feature	p-value	AIGR	Feature	p-value	AIGR
F1	0.0024	0.223	F15	>0.05	0.691
F2	>0.05	0.25	F16	>0.05	0.2645
F3	0.0312	0.018	F17	>0.05	0.2669
F4	0.0433	0.230	F18	0.0168	0.3147
F5	>0.05	0.1819	F19	>0.05	0.2513
F6	>0.05	0.0331	F20	>0.05	0.4230
F7	0.0011	0.2057	F21	0.0263	0.3128
F8	>0.05	0.079	F22	>0.05	0.3295
F9	0.0007	0.080	F23	>0.05	0.3975
F10	0.0001	0.038	F24	>0.05	0.4852
F11	>0.05	0.223	F25	>0.05	0.1641
F12	0.0003	0.168	F26	0.0001	0.1557
F13	0.0125	0.2411	F27	0.0001	0.2994
F14	0.0301	0.1697			

vertical segment and whole image CIN labels are from the same expert for classifier training and testing, denoted as RZ-RZ and SF-SF for the two experts, respectively. Individual vertical segment exact class classification results are given in Table 2. From Table 2, the highest individual classification accuracy (81.29%/80.10% based on labels from RZ-RZ/SF-SF, respectively) for exact classification was obtained using the Logistic classifier based on the Image Label approach for assigning CIN class labels to the individual vertical segments (all vertical segments within an image are assigned the image CIN label). Accuracies of 62.71%/64.92% (RZ-RZ/SF-SF) and 60.16%/63.23% (RZ-RZ/SF-SF) were obtained using the SVM and LDA classifiers, respectively, based on the Image Label approach for individual vertical segment labeling, which were used in [2](Note that "RZ-RZ" means that RZ's labels were used for both training and testing in the referenced experiment; likewise for SF).

The second set of experiments examined the impact of feature selection on CIN classification accuracy for the individual vertical segments. For feature evaluation and selection experiments, all 27 features extracted from the individual vertical segments with CIN truth labels from RZ were used as inputs to the SAS MLR algorithm as well as the

feature selector in $Weka$. A value of α = 0.05 was used to determine statistical significance for the input features for the SAS MLR. The $Weka$ feature selector ranks the features by an "attribute information gain ratio" (AIGR) which ranges from 0 to 1, with larger values indicating greater significance for the feature. The overall twenty-seven features with P values are presented in Table 3.

Based on the statistical significance of all the 27 features, the feature set selected using α = 0.05 consisted of F1, F3, F4, F7, F9, F10, F12, F13, F14, F18, F21, F22, F23, and F24. Note that all these features were selected based on the SAS MLR test of statistical significance except for F22, F23, and F24, which were selected since they have a relatively high information gain ratio (AIGR) among the 27 features [from 2nd place to 4th place in Table 3][2]. We compared discrimination accuracies using this reduced set of features to the results using the entire 27-feature set for fusion-based whole image classification based on (Section III) for combining the individual vertical segment classifications. Individual vertical segment classifications were generated using the SVM, LDA, Logistic Regression, and Random Forest classifiers based on the Image Label, Major Sub, and Image Sub approaches for obtaining individual vertical segment CIN labels for classifier training. For these experiments, the training and testing CIN labels were from the same expert, denoted as RZ-RZ and SF-SF, respectively. Exact class label and normal versus CIN classification whole image results are reported for the different classifiers based on all 27 features in Table 4 and the reduced feature set in Table 5.

In Table 6, the best confusion matrix result obtained using RZ-RZ labels for training-testing for the reduced feature set is shown, with an exact class label classification of 88.98% and normal versus CIN classification of 94.92%. Our highest previous results[4] for a 61 image dataset were 88.5% (exact classification accuracy) and 95.1% (normal vs. CIN) using the LDA classifier and RZ-RZ training-testing labels. For comparison purposes, Table 7

Table 4. Fusion-based whole image percentage correct cervical intraepithelial neoplasia discrimination rates using all features using the same expert for training and testing sets

	Classification scheme	SVM (RZ-RZ/ SF-SF)	LDA (RZ-RZ/ SF-SF)	Logistic (RZ-RZ/ SF-SF)	Random tree (RZ-RZ/ SF-SF)
Image	Exact	73.31/74.83	76.02/79.57	86.44/85.51	79.66/79.66
Label	Normal vs. CIN	87.29/88.98	84.75/85.59	94.07/93.22	88.14/88.98
Image	Exact	78.38/76.95	79.16/79.57	83.64/80.10	80.52/79.49
sub	Normal vs. CIN	91.53/92.37	93.22/92.37	96.61/91.53	90.68/89.83
Major	Exact	78.38/79.83	76.95/79.66	82.71/81.69	76.29/78.14
sub	Normal vs. CIN	84.75/86.44	87.29/87.29	94.07/94.92	84.75/86.44

Table 5. Fusion-based whole Image percentage correct cervical intraepithelial neoplasia discrimination rates using reduced features with the same expert for training and testing sets

	Classification scheme	SVM (RZ-RZ/ SF-SF)	LDA (RZ-RZ/ SF-SF)	Logistic (RZ-RZ/ SF-SF)	Random tree (RZ-RZ/ SF-SF)
Image	Exact	75.42/76.27	75.42/74.58	88.98/84.75	80.51/80.51
Label	Normal vs. CIN	82.20/83.05	81.36/82.20	94.92/92.37	90.68/92.37
Image	Exact	75.42/73.73	72.88/72.03	83.05/82.20	79.66/81.36
sub	Normal vs. CIN	84.75/88.14	77.12/78.81	91.53/90.68	87.29/89.83
Major	Exact	73.73/72.88	76.27/71.19	81.36/83.90	80.51/80.51
sub	Normal vs. CIN	85.59/83.90	83.05/81.36	91.53/92.37	87.29/88.14

presents the best confusion matrix result using RZ-RZ for training-testing for all 27 features on the 118 image set, which gives an exact class label classification of 86.44% and normal versus CIN classification of 94.92%.

4.2. Analysis of Results. In this section, we analyze the classification results from Section IVA, in four different ways: (a) a performance comparison among the classifiers (SVM, LDA, logistic, random forest) and (b) a performance comparison between previous research[4] and this study, (c) the impact on performance using intra-and inter-pathologist CIN truth labels for the classifier training and testing sets, and (d) a performance comparison

Table 6. Best confusion matrix results for fusion-based whole image classification using reduced feature set

	Expert RZ-RZ: Logistic			
	Normal (40)	CIN1 (25)	CIN2 (24)	CIN3 (29)
Normal	36	2	0	0
CIN1	2	22	3	0
CIN2	2	1	21	3
CIN3	0	0	0	26

Table 7. Best confusion matrix results for fusion-based whole image classification using all 27 features

	Expert RZ-RZ: Logistic			
	Normal (40)	CIN1 (25)	CIN2 (24)	CIN3 (29)
Normal	35	2	0	0
CIN1	2	22	2	1
CIN2	3	1	20	3
CIN3	0	0	2	25

between our classification results and the baseline results from the pathologists. The correct recognition rates for all classifiers investigated are presented using training-testing labels from RZ-SF/SF-RZ for all 27 features [Table 8] and the reduced feature set [Table 8].

From Tables 5 and 9, the logistic classifier exact class experiments for corresponding truth labels (Image Label, Image Sub, Major Sub) when reduced features are employed as the input feature vectors. The logistic classifier yielded a maximum improvement of 13.56% (75.42% from SVM and LDA to 88.98% for RZ-RZ) when using the truth tables from a single pathologist and a maximum improvement of 10.76% (71.95% from SVM to 82.71% for RZ-SF) using inter-pathologist truth tables as training and testing labels. In addressing with the unknown segments labeled as "0" by the pathologists, the labeling methods of Image Label and Major Sub had an impact on the overall classification results; the classification accuracies are improved when using the same classifiers but different labeling methods than the ones in previous research[2].

Table 8. Fusion-based whole image normal versus cervical intraepithelial neoplasia and exact cervical intraepithelial neoplasia discrimination rates using all 27 features (F1-F27) with expert training-testing labels of RZ-SF and SF-RZ

	Classification scheme	SVM (RZ-RZ/ SF-SF)	LDA (RZ-RZ/ SF-SF)	Logistic (RZ-RZ/ SF-SF)	Random tree (RZ-RZ/ SF-SF)
Image	Exact	72.88/71.95	72.88/72.88	81.36/78.81	72.88/71.19
Label	Normal vs. CIN	86.44/83.9	84.75/83.9	94.92/91.52	83.9/83.05
Image	Exact	75.42/72.88	75.42/71.19	78.81/77.29	75.42/76.19
sub	Normal vs. CIN	87.29/84.75	88.14/88.14	83.9/83.05	86.44/83.9
Major	Exact	72.88/71.19	72.88/72.88	77.29/76.95	72.88/73.73
sub	Normal vs. CIN	83.05/84.75	81.36/82.20	82.2//84.75	81.36/82.2

Table 9. Fusion-based whole image normal versus cervical intraepithelial neoplasia and exact cervical intraepithelial neoplasia discrimination rates using reduced features with training-testing labels of RZ-SF and SF-RZ

	Classification scheme	SVM (RZ-RZ/ SF-SF)	LDA (RZ-RZ/ SF-SF)	Logistic (RZ-RZ/ SF-SF)	Random tree (RZ-RZ/ SF-SF)
Image	Exact	71.95/72.45	72.88/75.42	82.71/78.39	75.42/75.42
Label	Normal vs. CIN	83.05/84.75	83.9/87.29	90.68/88.14	82.2/84.75
Image	Exact	75.42/74.58	74.58/73.73	76.95/77.29	75.42/76.19
sub	Normal vs. CIN	86.44/85.59	88.14/88.14	87.29/88.98	83.9/85.59
Major	Exact	72.88/71.19	72.88/72.88	77.29/76.95	72.88/73.73
sub	Normal vs. CIN	84.75/85.59	83.9/85.59	81.36//87.29	80.51/82.20

From the classification results for individual segment classification presented in Table 2 of Section IV, the logistic classifier gave an improvement of 10.19% (71.10% from LDA to 81.29% for RZ-RZ) and 8.49% (71.61% from LDA to 80.10% for SF-SF). Among all the results generated by the classifiers in this study, the highest individual segment classification accuracy is obtained with the logistic classifier, with the correct recognition rate of 81.29%. Compared with the accuracy obtained by the classifiers used in the previous research [2] SVM/LDA, the highest accuracy for the individual segment classification is 71.61%. An improvement of 9.32% is obtained by using logistic classifier. For fusion-based

whole image classification using the complete feature set (27 features), shown in Table 4 of Section IV, a decrease of 0.44% (88.5% LDA [2] to 86.44% logistic in this study) is obtained as the exact class image classification accuracy. A decrease of 2.6% (from 96.7% [2]LDA to 94.10% logistic in this study) is obtained as normal versus CIN correct rate. For the epithelium classification results using the reduced feature set shown in Table 5, a minimum improvement of 3.73% (from 85.25%[2] for LDA classifier to 88.98% in this study) is found. It can be observed that some of the classification accuracies drop when using one expert label as training and the other one as testing, compared with the results in Tables 8 and 9. For the logistic classifier, the highest exact classification rate was 88.98% [105/118 in Table 5] which was higher than 82.71% using one expert's labels for training (RZ) and the second expert's labels (SF) for testing [Table 9].

In examining the performance of our classification results, we also use the pathologists' truth labels of epithelium images to generate a baseline for exact classification accuracy. As shown in Table 10, the confusion matrix is obtained by fusing the pathologist truth labels of individual labels with the same fusion techniques of voting scheme which has already been explained in Section II. Note that for the individual vertical segment labels fusion; only the 61 images dataset is utilized to remain the study consistent with the previous study [2]. Table 10 highlights the variation in CIN grading for the expert pathologists for a 61 image data set, which differs from the 118 digitized histology image set used in this study. From Table 10, the experts RZ and SF had an exact class agreement in 78.7% (48/61) of the epithelium images. The experts differed by one CIN grade on the remaining 13 images (off-by-one). The exact class label fusion-based CIN discrimination results obtained in this study are comparable to the 78.7% expert agreement rate. The exact class LDA classifier result of 76.02% from Table 4 based on the training-testing CIN labels from RZ (denoted in this study as RZ-RZ) is based on the benchmark approach from the study by Guo et al. [2], where 88.5% is the exact class correct classification rate based on a 61 image data set from the study by Guo et al [2]. It should be noted that the 118 digitized histology image

set used in this research is a different data set than the 61 images from [2]. Consequently, the exact class discrimination rate of 76.02% provides the benchmark for comparing results in this study. The logistic regression classifier for the 118 image set yielded exact class discrimination results as high as 88.98%/85.51% (RZ-RZ/SF-SF) using the same expert for training-testing CIN labels and the image CIN label for each individual vertical segment, a 12.96%/5.94% (RZ-RZ/SF-SF) improvement for single expert over the baseline method[2]. The logistic regression method gave the highest vertical segment classification rate of 81.29%/80.10% (RZ-RZ/SF-SF), which fueled the higher fusion-based image classification. Overall, the CIN classification rates tended to be higher using the training-testing labels for the same expert than for training labels from one expert and testing labels from the other expert. Based on the logistic classifier, the same expert exact label results were 88.98%/85.51% (RZ-RZ/SF-SF) compared to training labels from one expert and testing labels from the other expert 82.71%/78.39% (RZ-SF/SF-RZ), an increase of 6.27%/7.12%, respectively. This result can be used to highlight the impact of building larger data sets where different experts are involved in truthing or diagnostically assessing parts of the data set.

Table 10. Confusion matrix classification baseline obtained from pathologist ground truth labels

	Fusion-based classfication (RZ/SF)			
	Normal (16/14)	CIN1 (13/14)	CIN2 (14/17)	CIN3 (18/16)
Normal	15/10	0/0	0	0
CIN1	1/4	13/13	2/3	0/0
CIN2	0/0	0/1	12/14	1/3
CIN3	0/0	0/0	0/0	17/13

For the logistic and random forest classifiers, which performed better in this study, it appears that using the same CIN label for each vertical segment in training and testing the different classifiers compared to using the local, individual expert determined CIN labels for training and testing the different classifiers resulted in slightly higher overall exact label

discrimination rates; there does not appear to be a corresponding trend in the exact label classification rates for the SVM and LDA classifiers. Guo et al[2]. reported the image-based exact label discrimination rates were much lower than the fusion-based voting of the individual vertical segment exact label classifications. It appears that the local CIN information from the individual vertical segments contributes to enhanced image-based exact label discrimination. However, variations in the vertical segment CIN truthing for an image do not appear to provide an improvement to an overall image CIN assessment.

The confusion matrix classification results presented in Table 9 show that by fusing the pathologists' labels without any prediction from classifiers, RZ's labels give an exact classification accuracy of 93.44% (57/61) and SF's labels indicates an exact correct recognition rate of 81.97% (50/61). Moreover, from the exact classification accuracy, we obtained in this study, the highest result of 88.98% falls in the range of this baseline provided from those two pathologists.

Table 11 presents a summary of the highest CIN classification results determined from this study for the different classifiers and training-testing expert truth label combinations and the highest classification results obtained from the experiments performed[2]. From Table 11, the exact class label results for the 118 image set examined in this study are comparable to the results reported for the 61 image set[2] based on all 27 features and the reduced feature set. Individual vertical segment results were not reported[2]. However, applying the same LDA classifier[2] to individual vertical segment classification from the 118 image set in this study showed an improvement of 16.87% (from LDA classifier 60.16%/63.23% (RZ-RZ/SF-SF)) to logistic regression 81.29%/80.10% from [Table 2]). In addition, comparing the LDA approach from[2] for fusion-based image classifier for the 118 image set yielded an improvement of 13.56%/10.17% with the logistic regression classifier (from LDA classifier 75.42%/74.58% [RZ-RZ/SF-SF] to logistic regression 88.98%/84.75% from [Table 5]) using the reduced feature set, and an improvement of 12.96/5.94% (from LDA classifier 76.02%/79.57% [RZ-RZ/SF-SF]) to logistic regression

88.98%/85.51% from [Table 4]) using all 27 features. Since exact class label is the most stringent of the scoring schemes we used, we interpret these results as showing a substantial gain in classification accuracy when using the logistic regression classifier for the extended image dataset of 118 histological images over the approaches explored in previous research[2].

Table 11. Summary of best classification accuracies: Current study versus previous research versus current

	LDA [2] (with 61 images): RZ-RZ /SF-SF	Current study: RZ-RZ /SF-SF	Current study: RZ train -SF test	Current study: SF train -RZ test
Fusion-based classification using 27 features (%)				
Exact	88.5[a]/82.0[a]	86.44/85.51[a]	81.36[a]	78.81[a]
Normal vs. CIN	96.7[a]/90.2[a]	96.612[b]/94.92[c]	94.92[a]	91.52[a]
Individual segment classification (%)				
Exact	Not reported	81.29[a]/80.10[a]		
Fusion-based classification using reduced features (%)				
Exact	88.52[a]/85.3[a]	88.98[a]/84.75[a]	82.81[a]	81.36[c]

Individual vertical segment labeling approach: [a]Image label, [b]Image sub, [c]Major sub. CIN: Cervical intraepithelial neoplasia

In examining the classification results, the majority of the exact class label classification errors are off-by-one CIN grade. Figure 5 shows an example of an image with expert label of CIN2 (RZ) that was labeled as a CIN3 by the LDA classifier.

From the basal membrane near the top of the epithelium in Figure 5 across the epithelium (downward toward the bottom), the nuclei distribution is relatively uniform in certain regions. The nuclear features, as well as the layer-by-layer Delaunay triangle features, highlight the relatively uniform distribution of nuclei in the vertical segments containing those regions, which correspond to a higher CIN grade. In other regions of the epithelium, the nuclei density is not as uniform across the epithelium, which could provide for a less severe CIN grade label for the epithelium. Figure 6 shows an example of an image with pathologist label of CIN2 (RZ) that was labeled as a CIN1 by the logistic classifier. This image has the texture and nuclei distribution which is more consistent with a CIN2 grade. However, the relative small nuclei area and lower color luminance in the epithelium leads to a lower CIN grade misclassification.

The overall algorithm was found to be robust in successful identification of nuclei. To evaluate nuclei detection, we manually counted nuclei in the two lightest-stained slides and the two darkest-stained slides. An average of 89.2% of the total number of nuclei in all four slides was detected. The 89.2% nuclei detection rate observed represents an advance over the results of Veta et al.[20], who detected nuclei at rates of 85.5%-87.5% (not strictly comparable, as these results were for breast cancer). The finding of a high percentage of nuclei in the lightest-and darkest-stained slides suggests that the algorithm is adaptable and robust with regard to varying staining.

The approach in this study expands the techniques of other studies that focus on the nucleus. We show in this work that the transition from normal to CIN3 affects the whole cell. We have shown that not only nuclei, but features of intercellular spaces are changed due to the more rapidly growing cells. Thus, one of the top four features by P value is the proportion of regions of cytoplasm in the image (F12).

5. CONCLUSION

In this study, we extended a localized, fusion-based image analysis approach for CIN classification to 118 digitized histology images. Twenty-seven features were explored, including the layer-by-layer triangle features and the nuclei as well as acellular features, as developed in previous research.[4] We conducted CIN discrimination experiments based on CIN truthing of the 118 image set by two pathologists (RZ/SF), including: (1) SF's CIN labels as training labels and testing labels. (2) RZ's CIN labels as training labels and testing labels. (3) RZ's CIN labels as training labels and SF's labels as testing labels. (4) SF's CIN labels as training labels and RZ's labels as testing labels. The vertical segments were classified using logistic regression, SVM, or LDA classifier, based on one of the four ways of labeled training data mentioned with a leave-one-out approach. We used a voting scheme to fuse the vertical segment classifications into a classification of the whole epithelium image. We evaluated the classification results with three scoring schemes, and compared the classification differences by classifiers, by scoring schemes, and the classification results of this research as compared to our previous work[2]. Experimental results showed that the logistic and random tree classifiers outperformed the benchmark SVM and LDA classifiers[2]. The logistic regression classifier gave exact class discrimination results as high as 88.98%/85.51% (RZ/SF) using the same expert for training-testing CIN labels and the image CIN label for each individual vertical segment, which is a 13.56%/10.17% (RZ-RZ/SF-SF) improvement for single expert over the baseline method[2] using the reduced features. The CIN classification rates tended to be higher using the training-testing labels for the same expert than for training labels from one expert and testing labels from the other expert. The exact class label fusion-based CIN discrimination results obtained in this study are comparable to the exact class expert agreement rate.

6. ACKNOWLEDGMENTS

In addition, we gratefully acknowledge the medical expertise and collaboration of Dr. Mark Schiffman and Dr. Nicolas Wentzensen, both of the National Cancer Institute's Division of Cancer Epidemiology and Genetics.

7. FINANCIAL SUPPORT AND SPONSORSHIP

This research was supported (in part) by the Intramural Research Program of the National Institutes of Health, National Library of Medicine, and Lister Hill National Center for Biomedical Communications.

8. CONFLICTS OF INTEREST

There are no conflicts of interest.

REFERENCES

[1] J. R. Egner, "AJCC Cancer Staging Manual," *JAMA*, vol. 304, p. 1726, 10 2010.

[2] P. Guo, K. Banerjee, R. Stanley, R. Long, S. Antani, G. Thoma, R. Zuna, S. R. S. Frazier, R. H. Moss, and W. V. Stoecker, "Nuclei-Based Features for Uterine Cervical Cancer Histology Image Analysis with Fusion-based Classification," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, pp. 1595–1607, 10 2015.

[3] S. De, R. J. Stanley, C. Lu, R. Long, S. Antani, G. Thoma, and R. Zuna, "A fusion-based approach for uterine cervical cancer histology image classification," *Computerized Medical Imaging and Graphics*, vol. 37, no. 7-8, pp. 475–87, 2013.

[4] F. P. Preparata and M. I. Shamos, *Computational geometry: an introduction*, vol. 47. 1985.

[5] L. He, L. R. Long, S. Antani, and G. R. Thoma, "Histology image analysis for carcinoma detection and grading," *Computer Methods and Programs in Biomedicine*, vol. 107, no. 3, pp. 538–556, 2012.

[6] R. Gonzalez and R. Woods, *Digital image processing*. 2002.

[7] Y. Wang, D. Crookes, O. S. Eldin, S. Wang, P. Hamilton, and J. Diamond, "Assisted diagnosis of cervical intraepithelial neoplasia (CIN)," *IEEE Journal on Selected Topics in Signal Processing*, vol. 3, pp. 112–121, 2 2009.

[8] M. Mignotte, "Segmentation by fusion of histogram-based K-means clusters in different color spaces," *IEEE Transactions on Image Processing*, vol. 17, no. 5, pp. 780–787, 2008.

[9] P. Guo, *Cervical cancer histology image feature extraction and classification*. PhD thesis, Missouri University of Science and Technology, Rolla, Missouri, 2014.

[10] J. Van Der Marel, W. G. V. Quint, M. Schiffman, M. M. van de Sandt, R. E. Zuna, S. T. Dunn, K. Smith, C. A. Mathews, M. A. Gold, J. Walker, and N. Wentzensen, "Molecular mapping of high-grade cervical intraepithelial neoplasia shows etiological dominance of HPV16," *International Journal of Cancer*, vol. 131, no. 6, pp. 946–53, 2012.

[11] C.-C. Chang and C.-j. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, pp. 1–39, 2011.

[12] S. L. Cessie and J. C. V. Houwelingen, "Ridge Estimators in Logistic Regression," *Applied Statistics*, vol. 41, no. 1, p. 191, 1992.

[13] M. A. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *SIGKDD Explorations*, vol. 11, no. 1, pp. 10–18, 2009.

[14] L. Breiman, "Manual on setting up, using, and understanding random forests v3. 1," *Technical Report, http://oz.berkeley.edu/users/breiman, Statistics Department University of California Berkeley, . . .* , p. 29, 2002.

[15] L. Breiman and A. Cutler, "Breiman and Cutler's random forests for classification and regression," *Package 'randomForest'*, p. 29, 2012.

[16] A. Agresti, *An Introduction to Categorical Data Analysis*. 2007.

[17] M. Pal, "Multinomial logistic regression-based feature selection for hyperspectral data," *International Journal of Applied Earth Observation and Geoinformation*, vol. 14, no. 1, pp. 214–220, 2012.

[18] T. Li, S. Zhu, and M. Ogihara, "Using discriminant analysis for multi-class classification: An experimental investigation," *Knowledge and Information Systems*, vol. 10, no. 4, pp. 453–472, 2006.

[19] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression 2nd ed. New York John Willey & Sons*. New York: John Willey & Sons, 2000.

[20] M. Veta, P. J. van Diest, R. Kornegoor, A. Huisman, M. A. Viergever, and J. P. W. Pluim, "Automatic Nuclei Segmentation in H&E Stained Breast Cancer Histopathology Images," *PloS one*, vol. 8, no. 7, p. e70221, 2013.

III. FEATURES ADVANCES TO AUTOMATICALLY FIND IMAGES FOR APPLICATION TO CLINICAL DECISION SUPPORT

P. Guo[1] H. Almubarak[1] R.Joe Stanley[1] Rodney Long[2] Sameer Antani[2] George Thoma[2]

Rosemary Zuna[3] Shelliane R. Frazier[4] Randy H. Moss[1] William V.Stoecker[5]

[1]Department of Electrical & Computer Engineering

Missouri University of Science and Technology, Rolla

[2]Lister Hill National Center for Biomedical Communications,

National Library of Medicine, National Institutes of Health, DHHS, Bethesda, MD

[3]Department of Pathology,

University of Oklahoma Health Sciences Center, Oklahoma City, OK

[4]Surgical Pathology Department,

University of Missouri Hospitals and Clinics, Columbia, MO

[4]Stoecker & Associates, Rolla, MO, USA

Email: stanleyj@mst.edu

ABSTRACT

Filtering through ever increasing sources of information to find relevant information for clinical decisions is a challenging task for clinicians. In biomedical publications, there are a variety of items that can provide evidence to aid the decision making process. One example is illustration image analysis and classification, which has been used to characterize and distinguish specific image modalities; this capability in turn has been used to assist in the evidence gathering process. This paper examines clinical decision support applications and extends previous research for illustration modality discrimination analysis.

Specifically, global, HSV histogram-based, and Gabor filter-based features are compared to histogram-based features for modality classification on a set of 12,056 images from 2004-2006 biomedical publication issues of Radiology and Radio Graphics that were manually annotated by modality (radiological, photo, etc.). Using a nearest neighbor classifier, average modality discrimination results were obtained as high as 99.98% using correlated features computed from Gabor filter spectral coefficients. These experimental results indicate that image features, particularly correlation-based features, can provide modality discrimination useful for clinical decision support applications.

1. INTRODUCTION

Clinical decision support (CDS) computer applications can potentially give healthcare professionals, patients, and researchers useful knowledge to improve healthcare and health related decisions. Considering the large and ever growing repositories of biomedical data, there is a demand for systems and tools to aid in finding useful information in biomedical publications, text databases, image databases, electronic health care records, clinical notes, and other sources, including full text, to support clinical decisions. The role of images in providing information for CDS is examined in this study, where an "image" can refer to visual materials in electronic healthcare records, databases, and articles in biomedical publications. Biomedical images include conventional images (MRI, CT, PET, for example), as well as illustrations, charts, and graphs. By moving beyond conventional text-based searching to combining both text and image features ("visual features") in search queries, the overall research goal is to enhance information retrieval from these entities for clinical decision support. The approach and the tools investigated take advantage of advances in Information Retrieval (IR), Content-based Image Retrieval (CBIR), and Natural Language Processing (NLP).

This research has focused on improving information retrieval of visual content from biomedical publications, in particular, by using features of the images themselves in combination with cues from text associated with the images. This includes using text from figure captions, image modality information from visual features and accompanying text [1, 2, 3], and annotation markers, such as arrows [4], letters or symbols embedded in images [2].

From the CDS perspective, knowing and differentiating image modality can impact an image's utility and improve the relevance of query results. Some previous document retrieval work has used the UMLS [5] term and concept query expansion engine in combination with fields from search results such as $MEDLINE$ citations (e.g., titles, abstracts and MeSH terms) and image features. This combination of attributes has been used to develop "visual keywords" with the goal of approximating image semantic labels [6, 7]. Automatic illustration identification has been explored for illustrations in medical publications which may assist a clinician in determining the usefulness of a particular publication for patient monitoring and treatment [1, 8, 9, 10].

A number of image features and techniques potentially useful for CDS have been applied in the field of Content-Based Image Retrieval (CBIR), including: 1) features of color, shape, and texture, and distance measures to compute similarity between images [9, 10, 11, 12, 13]; 2) Hough transform shape detection for region of interest determination and segmentation (has been used for lung images) [10]; 3) color analysis of stains for region of interest labeling (has been used for malaria cell images) [14]; 4) connecting the user and the database through a search engine with a feedback neural network architecture [15]; 5) query system modeling human interaction [16]; 6) Big Data use with query forms [17]; 7) use of image "key points" to identify salient parts of an image [18]; 7) combining image and text information for matrix similarity assessment [19]; 8) three-dimensional image analysis [8]; 9) latent topic models for computing image similarity [20]; 10) statistical model-

based image feature extraction using the wavelet domain and a Kullback divergence-based similarity measure for CBIR [21]; and 11) localized texture characterizations for CBIR for remote sensing applications [22].

There have also been numerous studies which explore the use of text information in for CDS applications, including: 1) extreme learning machine (ELM) and online sequential extreme learning machine (OSELM) with cuckoo search [23]; 2) cluster-based external expansion modelling with feedback [24]; 3) processing patient health record databases for matching, retrieval, and identification using templates for similarity assessments [25, 26, 27, 28, 29]; 4) graph theory and neural networks for literature mining [30]; 5) hash-based similarity searching [31]]; 6) fusion of image descriptors and text for medical image retrieval [32]; and 7) automatically supplementing references with images from articles for evidence finding [33]; and 8) demonstrating that image and text can yield retrieval accuracy appropriate for clinical evidence [33].

Recent publications related to the use of biomedical images in clinical decision support include: 1) an overview by Agarwal [34] of the critical steps part of computer-assisted detection (CAD) and computer-assisted diagnosis (CADx) systems: preprocessing, segmentation, region of interest (ROI) analysis, and assessment of detected structures and linear discriminant analysis (LDA) and support vector machine (SVM) approaches for these types of classification applications; 2) the use of image capture with mobile phone camera technology for cervical cancer screening in low resource parts of Africa [35]; and 3) an approach for vertebral level localization in spine radiographs as decision support for target localization in spine surgery [36].

This study builds off research related to biomedical image retrieval in the literature or related to image classification expected to be useful as preprocessing in clinical decision support systems, including: 1) automatic classification in a hierarchical taxonomy of figures from the biomedical literature [37]; 2) creating a comprehensive "visual ontology" for images in the biomedical literature [38]; 3) image modality classification, separation of

compound figures, and image retrieval using the 2013 ImageCLEF image set [39] (see [40] for an overview of the results of all ImageCLEF biomedical image retrieval tasks 2004-2013); 4) biomedical image modality classification using image clustering with respect to specified features, expert labeling of the (relatively few) clusters, and image classification based on the cluster labels [41]; 5) modality classification of biomedical literature figures comparing the effectiveness of SVM classification using hand- crafted features versus a deep learning classifier [42]; 6) extracting endoscopic images from the biomedical literature [43] and distinguishing true endoscopic images from confounding images; 7) classification of radiological signs in abdominal CT images [44]; 8) classification of view (frontal or lateral) in chest X-ray images [45]; 9) classification of Visible Human biomedical images into body segment classes (head and neck, thorax, abdomen, pelvis, and lower limb) by image features [46]; 10) methods to exploit "pointers" (such as arrows) or labels (such as letters or numbers) embedded within biomedical images, for image analysis and retrieval [47]; 11) the use of text associated with biomedical images to enhance image modality classification [48] and retrieval [49, 50]; and 12) modality- based classification over a set of 742 images manually annotated by modality (such as radiological or photo) selected from the 2004- 2005 issues of the British Journal of Oral and Maxillofacial Surgery using global, histogram- based, texture image illustration features, and basis function luminance histogram correlation features computed from the annotated images [1].

In this paper, a CDS application is presented that extends the image feature development work from [1]. As modality classification indices, basis function features created from the HSV histogram and Gabor filter to correlation features computed from the luminance histogram are compared. These quantities are applied to a set of medical publication illustrations and modalities examined in previous research [33]. The remainder of the paper is organized as follows: 1) description of the features and feature groups investigated, 2) modality classification experiments performed, 3) results and discussion, and 4) conclusions.

2. METHODOLOGY

2.1. Data Set Examined. In this study, images in various modalities were examined from the *ImageCLEFMed* 2010 dataset [51] from 2004-2006 issues of Radiology and *RadioGraphics* biomedical publications; these images were previously investigated by Demner-fushman et al. in [33] for feature development and classification. Table 1 provides a description of the categories of truthed images in this dataset. From Table 1, there are 2470 positive id images for all of the categories and 9586 negative id images for all of the categories. Positive id images represent truthed images in the designated categories. Negative id images are images from other categories that are similar to the designated categories for comparison. Figure 1 presents an image example from each of the category numbers listed in Table 1.

Table 1. Features Advances to Automatically Find Images for Application to Clinical Decision Support

Category Label	No. Positive Id Images	No. Negative Id Images
Doppler ultrasound images	286	513
CT images with emphysema	68	860
knee x-ray images	112	786
mediastinal CT	291	571
abdominal CT images showing liver blood vessels	299	721
chest CT images showing micro nodules	59	697
x-ray images containing one or more fractures	105	727
CT liver abscess	59	775
MRI or CT of colonoscopy	236	601
photographs of tumours	320	640
images of muscle cells	79	778
images containing a Budd-Chiari malformation	74	708
gastrointestinal neoplasm	273	607
pulmonary embolism all modalities	209	602

2.2. Features and Feature Groups Investigated. In prior research [1], the method of correlating basis functions with the luminance histogram for an image was found to be effective for discriminating image modalities [4]. These basis function correlation features have been explored in dermatology imaging research to provide gray level distribution information for skin lesion discrimination [1]. In this study, the basis function correlation features are extended to include the HSV histogram, both smoothed and unsmoothed, and Gabor features, denoted as Groups 1, 2, and 3, respectively. The details for these feature calculations are presented in the following sections.

2.3. Features and Feature Groups Investigated. In prior research [1], the method of correlating basis functions with the luminance histogram for an image was found to be effective for discriminating image modalities [4]. These basis function correlation features have been explored in dermatology imaging research to provide gray level distribution information for skin lesion discrimination [1]. In this study, the basis function correlation features are extended to include the HSV histogram, both smoothed and unsmoothed, and Gabor features, denoted as Groups 1, 2, and 3, respectively. The details for these feature calculations are presented in the following sections.

2.3.1. HSV Histogram Correlation Features. Group 1 and Group 2 features are computed from unsmoothed, and smoothed, one- dimensional HSV histograms, respectively [20]. These features are computed as follows. Each pixel in the image contributes to the histogram weighted values of its hue 'H' and intensity 'V', based on its saturation 'S'. Hence, the histogram has two components, the 'color components' representing the contribution of hue from each pixel, and the 'gray component', representing the contribution of the intensity value at each pixel. The histogram retains the smoothness between the adjacent components and this allows us to perform a window based smoothing of the histogram. "Saturation projection" is used to determine the weights for hue and for intensity. The weight is dependent on saturation level s. The weight of hue component, w_h(s) w_s and the weight of intensity of component w_i(s) are computed using the equations [13]:

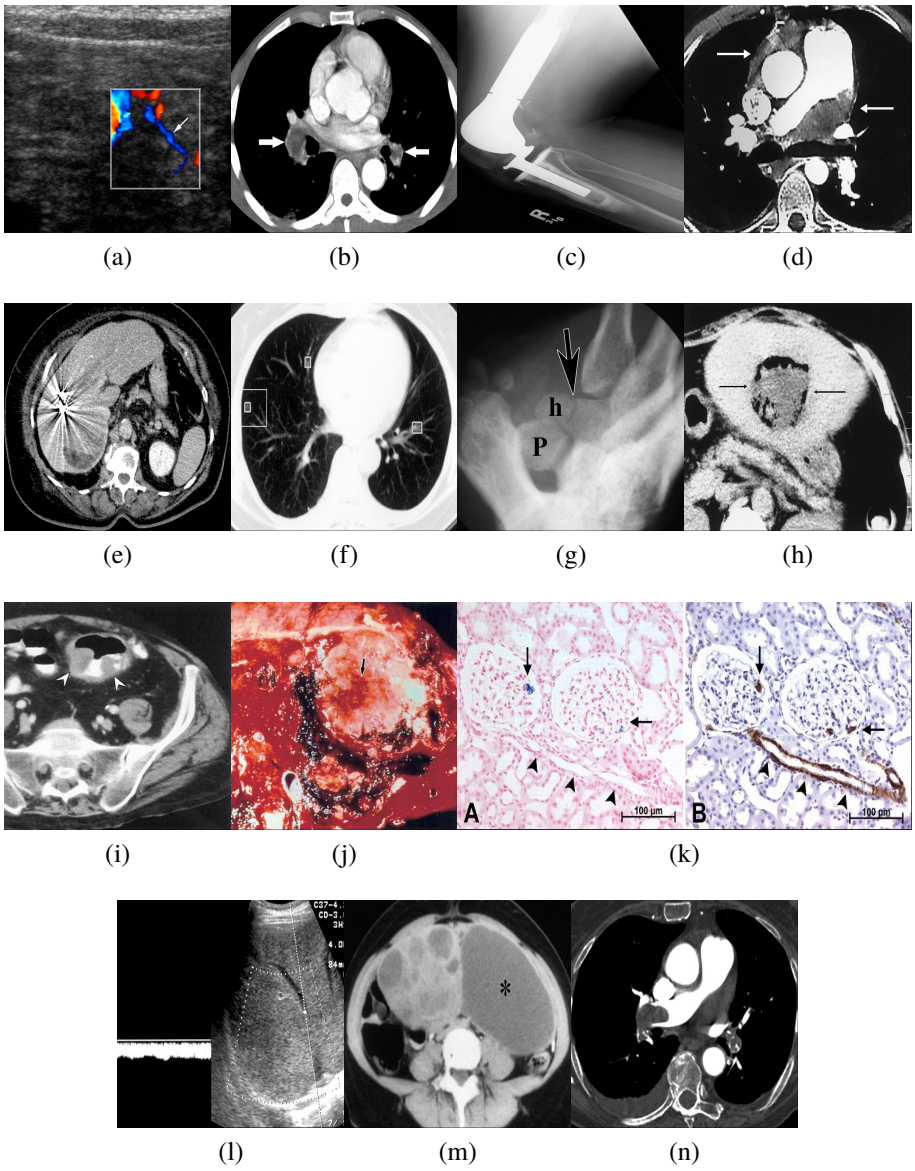

Figure 1. Image examples (positive id) from each of the category numbers listed in Table 1. (a) Doppler ultrasound image (reproduced with permission [52]). (b) CT image with emphysema (reproduced with permission [53]). (c) Knee x- ray image (reproduced with permission [54]). (d) Mediastinal CT image (reproduced with permission [55]). (e) Abdominal CT image showing liver blood vessels (reproduced with permission [56]). (f) Chest CT image showing micro nodules (reproduced with permission [57]). (g) X-ray image containing one or more fractures (reproduced with permission [58]). (h) CT liver abscess (reproduced with permission [59]). (i) MRI or CT of colonoscopy (reproduced with permission [60]). (j) Photograph of tumor (reproduced with permission [61]). (k) Image of muscle cells (reproduced with permission [62]). (l) Image containing a Budd-Chiari malformation (reproduced with permission [63]). (m) Gastrointestinal neoplasm (reproduced with permission [64]). (n) Pulmonary embolism all modalities (reproduced with permission [65]).

Table 2. Algorithm for generating HSV histogram.

For each pixel in image: Convert RGB values to HSV
Update histogram as follows: Hist[Round(H.MULT_FCTR)] = Hist[Round(H.MULT_FCTR)]+wh(s) Hist[Round(2πMULT_FCTR)+Round(V/DIV_FCTR)] =Hist[Round(2πMULT_FCTR)+Round(Hist[Round(V/DIV_FCTR)] +wi(s)

$$w_h(s) = s^\gamma \quad \text{where} \quad \gamma \in [0, 1] \tag{1}$$

$$w_i(s) = 1 - w_h(s) \tag{2}$$

The number of bins in the histogram is determined. Since the histogram consists of two components, the total number of bins is found by summing the number of color component bins and the number of gray component bins. Let N_h, N_g be the number of bins for the color and gray components, respectively, and let N be the total number of bins in the histogram [13]. Then

$$N_h = Round(2\pi MULT\_FCTR) + 1 \tag{3}$$

$$N_g = Round(I_{max}/DIV\_FCTR) + 1 \tag{4}$$

$$N = N_g + N_h \tag{5}$$

where: $MULT\_FCTR$: is the multiplying factor that determines the quantization level for the hues. I_{max} is the maximum intensity (generally 255). DIV_FCTR is a division factor that determines the number of quantized gray levels. The algorithm for generating the HSV histogram, denoted as Hist, is shown in Table 2 [13]:

Traditional histograms do not provide perceptual gradation of colors, but the HSV histogram retains this property. Thus, image-based features are explored based on the smoothed and unsmoothed histograms. The smoothing operation for the HSV histogram is given using the following equation [13]:

$$Hist_w(j) = \sum_{j+N}^{i=j-N} w(i-j)Hist(i) \tag{6}$$

where:

$j \in [0, N_h+N_h\text{-}1]$ and w(i-j) = $2^{-|i-j|}$

For the image HSV histogram, let $Hist_U$ and $Hist_S$ denote the unsmoothed and smoothed histograms, respectively. The basis function correlation features with the unsmoothed and smoothed HSV histograms are defined as follows. The basis function weighted density distribution (WDD) functions are given in Figure 2 below.

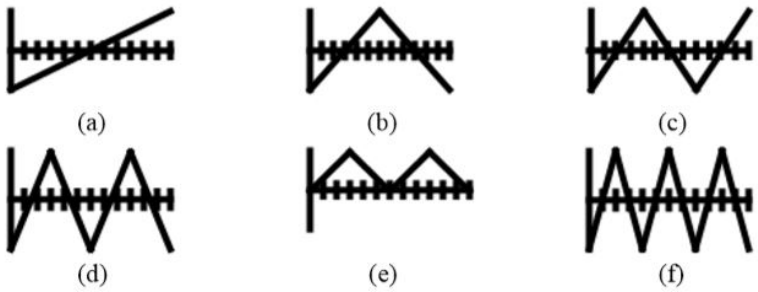

Figure 2. The WDD functions used for computing correlation-based features with the HSV unsmoothed and smoothed histograms and the Gabor filter coefficients (reproduced with permission [1])

The basis function WDD correlation features for a given image for unsmoothed HSV histogram are computed as:

$$Hist_{U,k}(j) = \sum_{N}^{i=1} Hist_U(i)W_k(i) \quad \text{for} \quad k = 1, 2, ..., 6 \tag{7}$$

$$Hist_{U,k}(j) = \sum_{N}^{i=1} \mid Hist_U(i) - Hist_U(i-1) \mid W_k(i) \quad \text{for} \quad k = 7, 8, ..., 12 \tag{8}$$

where $Hist_U(0) = 0$

The basis function WDD correlation features for the smoothed HSV histogram are similarly defined.

For each image, fifteen features are computed for the unsmoothed (Group 1) and smoothed (Group 2) HSV histograms. These features are: 1) the bin number which has the maximum count (mostFrequentComponent), 2) the average value of the color and of the gray components in the image ($avgVal$), 3) the standard deviation of the color and of the gray components in the image HSV histogram ($stdVal$), 4) 12 basis function features, denoted as $h_{HistU,1}$-$h_{HistU,12}$ for the unsmoothed HSV histogram features and $h_{HistS,1}$-$h_{HistS,12}$ for the smoothed HSV histogram features.

2.3.2. Group 3. The final set of features explored are based on Gabor filters. Gabor filters have been applied in CBIR for purposes such as extracting text regions from document images [66] and for texture analysis [12, 67] . The Gabor features are computed from an image using the following procedure. First, the image is resized to a square of dimensions *min_dim* x *min_dim*, where *min_dim* is the minimum of the row and column dimensions for the image. For example, a 512x768 image is resized to a 512x512 image. Second, if the image is color, it is normalized by applying a local luminance variance. Otherwise, the existing grayscale image is used. Third, the Gabor filter algorithm is applied to the resized image. The Gabor filter algorithm was used from [68] which is based on the algorithm presented in [68]. This algorithm uses orientations determined empirically at each scale of 8, 8, 4. Third, the array of spectral coefficients determined from the Gabor filter, denoted as I_{spect}, are correlated with the WDD basis functions to provide a profile of the spectral content and, thereby, texture information contained within the image. The WDD correlation features for a given image are defined as:

$$f_{I_{spect},k} = \sum_{192}^{i=1} I_{spect}(i) W_k(i) \quad \text{for} \quad k = 1, 2, ..., 6 \tag{9}$$

$$f_{I_{spect},k} = \sum_{192}^{i=1} \mid I_{spect}(i) - I_{spect}(i-1) \mid W_k(i) \quad \text{for} \quad k = 7, 8, ..., 12 \tag{10}$$

where $I_{spect}(0) = 0$ and 192 spectral coefficients found from the Gabor filter.

3. EXPERIMENTS PERFORMED

A benchmark technique based on different image features from [1] was used in order to evaluate these basis function-based feature groups (Group1-Group 3). These features are summarized in section 3.1. Two modality classification experiments were performed based on the positive and negative ID images within each of the 14 categories given in Table 1. In the first experiment, twenty randomly generated training and test sets were used for a nearest neighbor classifier (radial clustering algorithm with zero distance parameter, making it a nearest neighbor classifier) developed in [1]. Each training set consists of 90% of the image data feature vectors for each category, and each corresponding test set contains the remaining 10% of the image data feature vectors for each category. The image data feature vectors includes the features computed from Groups 1-3 and the benchmark features (see section 3.1) from [1]. In the second classification experiment, an image was classified into one of the 14 categories using the dataset for the positive id images within each category. Twenty randomly generated training and test sets were used. The training set consisted of 90% of the positive id images for each category, with the test set containing the remaining 10% of the positive id images for each category.

3.1. Benchmark Features. For image-based modality discrimination in [1], developed features were organized into three categories, as follows: 1) General Features, 2) Basis Function Features, and 3) Texture Features. The General Features quantified color, grayscale, histogram, and topology differences for grayscale and color figures. The General Features include: 1-3) standard deviation of red, green and blue values within the image, 4) percentage of the pixels in the image in which the green value is less than the red value and the green value is less than the blue value, 5) ratio of the pixels in the image with

luminance value greater than or equal to 250, 6) ratio of the number of pixels with the most frequently occurring luminance value to the area of the image, 7) square root of the area of the image, 8) ratio of the sum of the absolute differences between the red and green values and the red and blue values for each pixel in the image to the area of the image, 9) ratio of the pixels in the image with luminance value less than or equal to 30, 10) square root of the number of luminance histogram bins with counts greater than or equal to the scaled area of the image, 11) square root of the number of luminance histogram bins with counts greater than 0, 12) the square root of the number of luminance histogram bins with counts greater than or equal the scaled area of the image, and 13) estimate of the image fractal dimension. The Basis Function Features were the twelve WDD features computed from the image luminance histogram. The Texture Features were texture measures based on the Generalized Gray Level Spatial Dependence Models for Texture.

3.2. Nearest Neighbor Classifier. The nearest neighbor classifier was used as follows for the two experiments. For the first experiment, the minimum Euclidean distance for each test image was found for each positive and for each negative image in the training set, for each classification category. The test image feature vector is labeled a positive image for the category if the Euclidean distance to the minimum positive image is less than the Euclidean distance to the minimum negative image. Otherwise, the test image is labeled a negative image. These steps were applied to classifying each test image based on the feature vectors determined for each of the 14 categories given in Table 1, over 20 training and test sets. The average test results are reported over these 20 trials. For the second experiment, the minimum Euclidean distance was computed from each test image to each positive image in the training set, for all of the classification categories. The test image feature vector is assigned to the category of the positive image with the minimum Euclidean distance. These steps were applied to classifying each test image based on the feature vectors found from all of the categories given in Table 1 over 20 training and test sets. The average and standard deviation classification results were reported over these 20 trials.

4. EXPERIMENTAL RESULTS AND DISCUSSION

The experiments performed evaluated the modality discrimination capability of the proposed image-based features computed by correlating the basis functions (see Figure 2) with the unsmoothed HSV histogram (Group 1), smoothed HSV histogram (Group 2), and the Gabor filter spectral coefficient array (Group 3). Classification results using these feature groups are compared to benchmark color, grayscale, histogram, and topology features (General Features), basis function correlation features with the luminance histogram (Basis Function Features) and the Generalized Gray Level Spatial Dependence Models for Texture Features (Texture Features) from [1]. Modality classification experiments were performed on the 2470 positive id and 9586 negative id images for all fourteen modality categories shown in Table 1. Table 3 presents the test results using the nearest neighbor classifier for feature Groups 1-3 and the General Features, Basis Function Features, and Texture Features benchmarks from [1]. The test results for each of the 20 randomly generated training/test sets are given with the mean and standard deviation for the respective feature groups. Table 4 presents the average positive and negative nearest neighbor test classification results over 20 randomly generated training and test sets for all 14 categories (multi- class) for feature Groups 1-3 and the General Features, Basis Function Features, and Texture Features benchmarks from [1].

From Tables 3 and 4, there are several observations. First, from Table 3, all of the feature groups provided the capability to effectively distinguish the positive labeled images from the negative labeled images for each category. Second, from Table 3, the WDD- based features in feature groups Groups 1-3 and Basis Function Features [1] yielded slightly higher average classification rates than the other feature groups, which are based on global image and texture features. In these feature groups WDD correlation-based features were computed using the luminance image (Basis Function Features [1]), the HSV space (with and without HSV histogram smoothing (Groups 1 and 2, respectively)), and with the Gabor filter coefficients (Group 3). Group 1 features based on correlating the basis functions with

Table 3. Average test results using the nearest neighbor classifier for feature Groups 1-3 and benchmark features [3] for 20 training/test sets.

Category	General Features [1]	Basis Function Features [1]	Texture Features [1]	Group1	Group2	Group3
Dopplerultrasound images	91.13	99.75	96.75	99.88	84.50	99.88
CT images with emphysema	98.01	99.19	99.78	99.30	92.58	100.00
knee x-ray images	96.72	99.28	99.67	99.33	90.94	100.00
mediastinal CT	97.67	98.20	98.78	99.88	86.34	100.00
abdominal CT images showing liver blood vessels	90.83	99.95	99.12	99.51	86.62	100.00
chest CT images showing micro nodules	97.37	99.87	99.54	99.87	92.24	100.00
x-ray images containing one or more fractures	97.17	96.27	99.88	99.22	89.58	100.00
CT liver abscess	98.86	99.46	99.94	99.82	95.48	100.00
MRI or CT of colonoscopy	83.45	99.29	96.07	98.15	85.89	100.00
photographs of tumours	76.98	99.27	99.32	99.84	69.74	100.00
images of muscle cells	89.94	99.19	99.48	99.71	93.14	100.00
images containing a Budd-Chiari malformation	96.47	99.81	99.04	99.94	94.81	99.94
gastrointestinal neoplasm	72.90	94.66	80.68	91.36	74.49	99.83
pulmonary embolism all modalities	77.16	97.04	82.65	99.26	75.68	100.00
Average over all modality categories	90.33	98.66	96.48	98.93	86.57	99.98

Table 4. Multi-category test percentage correct results using nearest neighbor classifier for feature Groups 1-3 and benchmark features [3] for 20 training/test sets are presented with mean and standard deviation.

Training/ Test Set	General Features [1]	Basis Function Features [1]	Texture Features [1]	Group1	Group2	Group3
1	75.71	93.93	88.26	94.74	90.69	98.79
2	77.33	95.14	91.50	95.14	93.12	98.79
3	83.00	94.74	91.50	94.74	92.31	99.19
4	77.33	91.09	87.45	92.31	91.90	98.79
5	74.49	93.12	87.45	94.33	90.28	97.57
6	76.52	91.90	89.88	92.31	91.09	97.98
7	76.11	94.33	89.88	93.52	93.52	99.19
8	70.04	91.90	89.07	93.12	88.26	98.38
9	76.52	95.14	90.69	93.93	91.90	98.38
10	70.04	89.47	87.04	91.50	89.07	98.79
11	77.73	95.14	89.47	91.50	90.69	97.57
12	77.33	93.93	92.31	89.88	92.31	99.60
13	74.49	90.69	90.69	93.52	90.69	99.19
14	70.85	89.07	87.85	95.55	92.31	99.60
15	75.71	92.71	89.88	94.33	92.31	98.79
16	75.30	91.50	89.07	95.55	94.33	98.38
17	74.49	91.50	89.47	96.36	91.09	99.60
18	73.68	93.12	89.07	93.93	89.07	98.79
19	76.52	90.28	89.88	96.76	92.71	98.38
20	78.14	91.90	89.07	93.52	91.09	100.00
Mean	75.57	92.53	89.47	93.83	91.44	98.79
Std	2.98	1.90	1.43	1.73	1.54	0.66

unsmoothed HSV histogram (98.93%) outperformed the Group 2 features based on correlating the basis functions with the smoothed HSV histogram (86.57%). Overall, the basis function features based on correlating the WDD functions with the Gabor filter spectral coefficients yielded the highest modality classification results, with an average of 99.98% over the 14 different categories. This is a slight improvement of 1.32% over correlating the WDD functions with the luminance histogram (Basis Function Features [1]). Third, from Table 4, the WDD correlation-based features yielded the highest overall multi-category discrimination results. The Group 3 features, computed based on correlating the spectral coefficients from Gabor filtering with the WDD functions, produced the highest average classification results of 98.79%. These results and the other relatively high classification results based on the WDD correlation-based features, Basis Function Features [1] based on the luminance histogram (92.53%) and Group 1 based on the unsmoothed HSV histogram (93.83%) may indicate there is discrimination information in the distribution of gray levels and HSV values as well as the spectral coefficients over the images from the different categories in this data set. In the published literature, the WDD correlation-based features have been applied to extract symmetry and distribution information from histograms of malignant melanoma labeled colors for skin lesion discrimination [1]. This research appears to show another potential application for these basis function features in extracting distribution-based information which is similar for images of the different database categories. The nearest neighbor classifier results presented here highlight the similarity of images within each category using image-to-image matching. Fourth, the experimental results show that the General Features (global color and luminance features) from [1] and texture-based features (Generalized Gray Level Spatial Dependence Models) from [1] provided the lowest discrimination information for multi-category classification, with average test results of 75.57% and 89.47%, respectively. In the context of CDS applications, the image-based features presented in this study, particularly the basis function features from Group 3 which

correlate the array of Gabor filter spectral coefficients with the WDD functions can be used to discern image modalities that are representative of different types of biomedical information.

5. CONCLUSION

In this research, modality image classification was investigated using feature groups generated from HSV histograms and Gabor filters, and showed that these feature provide discrimination capability for positive/negative classification. For selected modality categories, the classification results show the potential for using image feature-based and machine learning classification in clinical decision support. Overall, the correlated WDD features with the spectral coefficients determined from the Gabor filtering achieved average classification as high as 99.98% for the experimental data set. The modality classification results of 98.93% obtained by correlating the WDD basis functions with the unsmoothed HSV histogram outperforms the feature groups from previous research [1]. Future research can be focused on principal feature analysis and finding the most significant features, and in down-sizing the current feature groups.

As an initial step in characterizing the visual content for biomedical information retrieval systems, positive results were achieved in image modality classification, and have improved on previous research.

6. ACKNOWLEDGEMENT

This work was supported by NLM under contract number 276200800413P and the Intramural Research Program of the National Institutes of Health (NIH), NLM, and Lister Hill National Center for Biomedical Communications (LHNCBC).

REFERENCES

[1] R. J. Stanley, S. De, D. Demner-Fushman, S. Antani, and G. R. Thoma, "An image feature-based approach to automatically find images for application to clinical decision support," *Computerized Medical Imaging and Graphics*, vol. 35, no. 5, pp. 365–372, 2011.

[2] S. De, R. J. Stanley, B. Cheng, S. Antani, R. Long, and G. Thoma, "Automated Text Detection and Recognition in Annotated Biomedical Publication Images," *Int. J. Healthc. Inf. Syst. Inform.*, vol. 9, pp. 34–63, 4 2014.

[3] B. Cheng, R. J. Stanley, S. De, S. Antani, and G. R. Thoma, "Automatic Detection of Arrow Annotation Overlays in Biomedical Images," *International Journal of Healthcare Information Systems and Informatics*, vol. 6, pp. 23–41, 34 2011.

[4] B. Cheng, R. Wang, S. Antani, R. J. Stanley, and G. R. Thoma, "Graphical image classification combining an evolutionary algorithm and binary particle swarm optimization," p. 829703, 1 2012.

[5] D. A. Lindberg, B. L. Humphreys, and A. T. McCray, "The Unified Medical Language System.," *Methods Archive*, vol. 32, no. 4, pp. 281–291, 1993.

[6] M. Rahman, S. K. Antani, G. R. Thoma, and U. S. N. Library, "A Medical Image Retrieval Framework in Correlation Enhanced Visual Concept Feature Space," pp. 6–9, 2009.

[7] M. M. Rahman, S. K. Antani, and G. R. Thoma, "A classification-driven similarity matching framework for retrieval of biomedical images," in *MIR 2010 - Proceedings of the 2010 ACM SIGMM International Conference on Multimedia Information Retrieval*, pp. 147–154, 2010.

[8] Y. Xue, T. Cheng, X. Xu, Z. Gao, Q. Li, X. Liu, X. Wang, R. Song, X. Ju, and Q. Zhang, "High-accuracy and real-time 3D positioning, tracking system for medical imaging applications based on 3D digital image correlation," *Optics and Lasers in Engineering*, vol. 88, pp. 82–90, 2017.

[9] S. Zhang and D. Metaxas, "Large-Scale Medical Image Analytics: Recent Methodologies, Applications and Future Directions," *Medical Image Analysis*, vol. 33, pp. 98–101, 2016.

[10] M. Sundarapandian, R. Kalpathi, R. A. C. Siochi, and A. S. Kadam, "Lung diaphragm tracking in CBCT images using spatio-temporal MRF," *Computerized Medical Imaging and Graphics*, vol. 53, pp. 9–18, 2016.

[11] T. J. Bright, A. Wong, R. Dhurjati, E. Bristow, L. Bastian, R. R. Coeytaux, G. Samsa, V. Hasselblad, J. w. Williams, M. D. Musty, L. Wing, A. S. Kendrick, G. D. Sanders, and D. Lobach, "Annals of Internal Medicine Review Effect of Clinical Decision-Support Systems," *Annals of Internal Medicine*, vol. 157, no. 1, pp. 29–43, 2012.

[12] D. Zhang, A. Wong, M. Indrawan, and G. Lu, "Content-based Image Retrieval Using Gabor Texture Features," *IEEE Transactions PAMI*, vol. 3656 LNCS, p. 13âĂŞ15, 2000.

[13] P. S. Hiremath and J. Pujari, "Content Based Image Retrieval based on Color , Texture and Shape features using Image and its complement," *International Journal of Computer Science and Security*, vol. 1, no. 4, pp. 25–35, 2007.

[14] S. Ghosh and A. Ghosh, "Content Based Retrieval of Malaria Positive Images from a Clinical Database VIA Recognition in RGB Colour Space," *Advances in Intelligent Systems and Computing*, vol. 248, 2014.

[15] P. H. Bugatti, D. S. Kaster, M. Ponciano-Silva, C. Traina, P. M. Azevedo-Marques, and A. J. M. Traina, "PRoSPer: Perceptual similarity queries in medical CBIR systems through user profiles," *Computers in Biology and Medicine*, vol. 45, no. 1, pp. 8–19, 2014.

[16] G. W. Hruby, L. V. Rasmussen, D. Hanauer, V. L. Patel, J. J. Cimino, and C. Weng, "A multi-site cognitive task analysis for biomedical query mediation," *International Journal of Medical Informatics*, vol. 93, pp. 74–84, 2016.

[17] S. Istephan and M. R. Siadat, "Unstructured medical image query using big data - An epilepsy case study," *Journal of Biomedical Informatics*, vol. 59, pp. 218–226, 2016.

[18] E. Tramontana, "Graphic object feature extraction system based on cuckoo search algorithm," 2016.

[19] M. Behrisch, B. Bach, M. Hund, M. Delz, L. von Ruden, J.-D. Fekete, and T. Schreck, "Magnostics: Image-based Search of Interesting Matrix Views for Guided Network Exploration," *IEEE Transactions on Visualization and Computer Graphics*, vol. 2626, no. c, pp. 1–1, 2016.

[20] F. Zhang, S. Member, Y. Song, and W. Cai, "Pairwise Latent Semantic Association for Similarity Computation in Medical Imaging," vol. 9294, no. c, pp. 1–12, 2015.

[21] H. Yuan and X.-p. Zhang, "Texture Image Retrieval Based on a Gaussian Mixture Model and Similarity Measure Using a Kullback Divergence," pp. 1867–1870, 2004.

[22] K. Jalaja, C. Bhagvati, B. L. Deekshatulu, and A. K. Pujari, "Texture Element Feature Characterizations for CBIR," vol. 500046, pp. 733–736, 2005.

[23] P. Mohapatra, S. Chakravarty, and P. K. Dash, "An improved cuckoo search based extreme learning machine for medical data classification," *Swarm and Evolutionary Computation*, vol. 24, pp. 25–49, 2015.

[24] H.-S. Oh and Y. Jung, "Cluster-based query expansion using external collections in medical information retrieval.," *Journal of biomedical informatics*, vol. 58, pp. 70–79, 2015.

[25] D. A. Szlosek and J. M. Ferretti, "Using Machine Learning and Natural Language Processing Algorithms to Automate the Evaluation of Clinical Decision Support in Electronic Medical Record Systems," no. August, 2016.

[26] M. Zalis and M. Harris, "Advanced Search of the Electronic Medical Record: Augmenting Safety and Efficiency in Radiology," *Journal of the American College of Radiology*, vol. 7, no. 8, pp. 625–633, 2010.

[27] L. Marco-Ruiz, C. Pedrinaci, J. Maldonado, L. Panziera, R. Chen, and J. G. Bellika, "Publication, discovery and interoperability of Clinical Decision Support Systems: A Linked Data approach," *Journal of Biomedical Informatics*, vol. 62, pp. 243–264, 2016.

[28] Y. Molenbruch, K. Braekers, A. Caris, and G. Vanden Berghe, "Multi-directional local search for a bi-objective dial-a-ride problem in patient transportation," *Computers & Operations Research*, vol. 77, pp. 58–71, 2017.

[29] A. M. Redd, A. V. Gundlapalli, G. Divita, M. E. Carter, L.-T. Tran, and M. H. Samore, "A Pilot Study of a Heuristic Algorithm for Novel Template Identification from {VA} Electronic Medical Record Text," *Journal of Biomedical Informatics*, pp. –, 2016.

[30] J. Jiang, J. Zheng, C. Zhao, J. Su, Y. Guan, and Q. Yu, "Clinical-decision support based on medical literature: A complex network approach," *Physica A: Statistical Mechanics and its Applications*, vol. 459, pp. 42–54, 2016.

[31] X. Liu, L. Huang, C. Deng, B. Lang, and D. Tao, "Query-Adaptive Hash Code Ranking for Large- Scale Multi-View Visual Search," vol. 25, no. 10, pp. 4514–4524, 2016.

[32] A. Mourão, F. Martins, and J. Magalhães, "Multimodal medical information retrieval with unsupervised rank fusion," *Computerized Medical Imaging and Graphics*, vol. 39, pp. 35–45, 2015.

[33] D. Demner-fushman, S. Antani, and G. R. Thoma, "Automatically Finding Images for Clinical Decision Support," pp. 139–144, 2007.

[34] J. P. Agrawal, B. J. Erickson, and C. E. Kahn, "Imaging Informatics: 25 Years of Progress," *IMIA Yearbook*, vol. 25, pp. S23–S31, 6 2016.

[35] C. Peterson, D. Rose, J. Mink, and D. Levitz, "Real-Time Monitoring and Evaluation of a Visual-Based Cervical Cancer Screening Program Using a Decision Support Job Aid," *Diagnostics*, vol. 6, p. 20, 5 2016.

[36] T. De Silva, S.-F. L. Lo, N. Aygun, D. M. Aghion, A. Boah, R. Petteys, A. Uneri, M. D. Ketcha, T. Yi, S. Vogt, G. Kleinszig, W. Wei, M. Weiten, X. Ye, A. Bydon, D. M. Sciubba, T. F. Witham, J.-P. Wolinsky, and J. H. Siewerdsen, "Utility of the LevelCheck Algorithm for Decision Support in Vertebral Localization," *SPINE*, vol. 41, pp. E1249–E1256, 10 2016.

[37] S. Antani, D. You, M. Simpson, M. Rahman, D. Demner-Fushman, and G. Thoma, "The Role of Image Modality and Visual Characteristics in Archiving Biomedical Images," in *Proc of IS&T Archiving Conference.*, Society for Imaging Science and Technology, 2013.

[38] M. Rahman, D. You, M. S. Simpson, S. K. Antani, D. Demner-Fushman, and G. R. Thoma, "An Interactive Image Retrieval Framework for Biomedical Articles Based on Visual Region-of- Interest (ROI) Identification and Classification," in *2012 IEEE Second International Conference on Healthcare Informatics, Imaging and Systems Biology*, pp. 50–50, IEEE, 9 2012.

[39] M. S. Simpson, D. You, M. M. Rahman, Z. Xue, D. Demner-Fushman, S. Antani, and G. Thoma, "Literature-based biomedical image classification and retrieval," *Computerized Medical Imaging and Graphics*, vol. 39, pp. 3–13, 1 2015.

[40] J. Kalpathy-Cramer, A. G. S. de Herrera, D. Demner-Fushman, S. Antani, S. Bedrick, and H. Müller, "Evaluating performance of biomedical image retrieval systems–an overview of the medical image retrieval task at ImageCLEF 2004-2013.," *Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society*, vol. 39, pp. 55–61, 1 2015.

[41] S. Vajda, D. You, S. K. Antani, and G. R. Thoma, "Label the many with a few: Semi-automatic medical image modality discovery in a large image collection," in *2014 IEEE Symposium on Computational Intelligence in Healthcare and e-health (CICARE)*, pp. 167–173, IEEE, 12 2014.

[42] Z. Xue, M. M. Rahman, S. Antani, L. R. Long, D. Demner-Fushman, and G. R. Thoma, "Modality Classification for Searching Figures in Biomedical Literature," in *2016 IEEE 29th International Symposium on Computer-Based Medical Systems (CBMS)*, pp. 152–157, IEEE, 6 2016.

[43] Z. Xue, D. You, S. Chachra, S. Antani, L. R. Long, D. Demner-Fushman, and G. R. Thoma, "Extraction of endoscopic images for biomedical figure classification [9418-23]," in *Proc. SPIE. 9418, Medical Imaging 2015: PACS and Imaging Informatics: Next Generation and Innovations, 94180P*, 2015.

[44] Z. Xue, D. You, S. Antani, L. R. Long, D. Demner-Fushman, and G. R. Thoma, "Classification of visual signs in abdominal CT image figures in biomedical literature," vol. 9039, p. 90390L, International Society for Optics and Photonics, 3 2014.

[45] Z. Xue, D. You, S. Candemir, S. Jaeger, S. Antani, L. R. Long, and G. R. Thoma, "Chest X-ray Image View Classification," in *2015 IEEE 28th International Symposium on Computer-Based Medical Systems*, pp. 66–71, IEEE, 6 2015.

[46] Z. Xue, S. Antani, L. R. Long, D. Demner-Fushman, and G. R. Thoma, "Body Segment Classification for Visible Human Cross Section Slices," in *2014 IEEE 27th International Symposium on Computer-Based Medical Systems*, pp. 199–204, IEEE, 5 2014.

[47] D. You, M. Simpson, S. Antani, D. Demner-Fushman, and G. R. Thoma, "A robust pointer segmentation in biomedical images toward building a visual ontology for biomedical article retrieval," vol. 8658, p. 86580Q, International Society for Optics and Photonics, 2 2013.

[48] D. You, M. M. Rahman, S. Antani, D. Demner-Fushman, and G. R. Thoma, "Text- and content-based biomedical image modality classification," vol. 8674, p. 86740L, International Society for Optics and Photonics, 3 2013.

[49] D. You, M. Simpson, S. Antani, D. Demner-Fushman, and G. R. Thoma, "Annotating image ROIs with text descriptions for multimodal biomedical document retrieval," vol. 8658, p. 86580D, International Society for Optics and Photonics, 2 2013.

[50] D. You, S. Antani, D. Demner-Fushman, and G. R. Thoma, "Does Figure-Text Improve Biomedical Article Retrieval? A Pilot Study," in *2014 IEEE 27th International Symposium on Computer-Based Medical Systems*, pp. 471–472, IEEE, 5 2014.

[51] "No Title."

[52] R. Faingold, A. Daneman, G. Tomlinson, P. S. Babyn, D. E. Manson, A. Mohanta, A. M. Moore, J. Hellmann, C. Smith, T. Gerstle, and J. H. Kim, "Necrotizing Enterocolitis: Assessment of Bowel Viability with Color Doppler US," *Radiology*, vol. 235, pp. 587–594, 5 2005.

[53] C. S. Restrepo, L. Diethelm, J. A. Lemos, E. Velásquez, T. A. Ovella, S. Martinez, J. Carrillo, and D. F. Lemos, "Cardiovascular Complications of Human Immunodeficiency Virus Infection," *RadioGraphics*, vol. 26, pp. 213–231, 1 2006.

[54] M. S. Taljanovic, T. B. Hunter, M. D. Miller, and J. E. Sheppard, "Gallery of Medical Devices," *RadioGraphics*, vol. 25, pp. 859–870, 5 2005.

[55] N. Hiller, S. Lieberman, T. Chajek-Shaul, J. Bar-Ziv, and D. Shaham, "Thoracic Manifestations of Behçet Disease at CT," *RadioGraphics*, vol. 24, pp. 801–808, 5 2004.

[56] H. P. Clark, W. F. Carson, P. V. Kavanagh, C. P. H. Ho, P. Shen, and R. J. Zagoria, "Staging and Current Treatment of Hepatocellular Carcinoma," *RadioGraphics*, vol. 25, pp. S3–S23, 10 2005.

[57] K. Awai, K. Murao, A. Ozawa, M. Komi, H. Hayakawa, S. Hori, and Y. Nishimura, "Pulmonary Nodules at Chest CT: Effect of Computer-aided Diagnosis on RadiologistsâĂŹ Detection Performance," *Radiology*, vol. 230, pp. 347–352, 2 2004.

[58] A. G. Blum, J.-P. Zabel, R. Kohlmann, T. Batch, K. Barbara, X. Zhu, G. Dautel, and F. Dap, "Pathologic Conditions of the Hypothenar Eminence: Evaluation with Multidetector CT and MR Imaging," *RadioGraphics*, vol. 26, pp. 1021–1044, 7 2006.

[59] H. Rhim, G. D. Dodd, K. N. Chintapalli, B. J. Wood, D. E. Dupuy, J. L. Hvizda, P. E. Sewell, and S. N. Goldberg, "Radiofrequency Thermal Ablation of Abdominal Tumors: Lessons Learned from Complications," *RadioGraphics*, vol. 24, pp. 41–52, 1 2004.

[60] P. J. Pickhardt, "Differential Diagnosis of Polypoid Lesions Seen at CT Colonography (Virtual Colonoscopy)," *RadioGraphics*, vol. 24, pp. 1535–1556, 11 2004.

[61] S. R. Prasad, H. Wang, H. Rosas, C. O. Menias, V. R. Narra, W. D. Middleton, and J. P. Heiken, "Fat-containing Lesions of the Liver: Radiologic-Pathologic Correlation," *RadioGraphics*, vol. 25, pp. 321–331, 3 2005.

[62] C. Bos, Y. Delmas, A. Desmoulière, A. Solanilla, O. Hauger, C. Grosset, I. Dubus, Z. Ivanovic, J. Rosenbaum, P. Charbord, C. Combe, J. W. M. Bulte, C. T. W. Moonen, J. Ripoche, and N. Grenier, "In Vivo MR Imaging of Intravascularly Injected Magnetically Labeled Mesenchymal Stem Cells in Rat Kidney and Liver," *Radiology*, vol. 233, pp. 781–789, 12 2004.

[63] T. Berrocal, M. Parrón, A. Álvarez-Luque, C. Prieto, and M. L. Santamaría, "Pediatric Liver Transplantation: A Pictorial Essay of Early and Late Complications," *RadioGraphics*, vol. 26, pp. 1187–1209, 7 2006.

[64] A. D. Levy, N. Patel, N. Dow, R. M. Abbott, M. Miettinen, and L. H. Sobin, "Abdominal Neoplasms in Patients with Neurofibromatosis Type 1: Radiologic-Pathologic Correlation," *RadioGraphics*, vol. 25, pp. 455–480, 3 2005.

[65] C. Wittram, M. M. Maher, A. J. Yoo, M. K. Kalra, J.-A. O. Shepard, and T. C. McLoud, "CT Angiography of Pulmonary Embolism: Diagnostic Criteria and Causes of Misdiagnosis," *RadioGraphics*, vol. 24, pp. 1219–1238, 9 2004.

[66] F. Nourbakhsh, P. B. Pati, and A. G. Ramakrishnan, "Text Localization and Extraction from Complex Gray Images," pp. 776–785, Springer, Berlin, Heidelberg, 2006.

[67] H. A. Jalab, "Image retrieval system based on color layout descriptor and Gabor filters," in *2011 IEEE Conference on Open Systems*, pp. 32–36, IEEE, 9 2011.

[68] A. Oliva and A. Torralba, "Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope," *International Journal of Computer Vision*, vol. 42, no. 3, pp. 145–175, 2001.

SECTION

2. SUMMARY AND CONCLUSIONS

This dissertation proposes data fusion techniques for uterine cervical cancer intraepithelial neoplasia (CIN) classification, image information retrieval for clinical decision support system . Data fusion was applied to these areas to extend the state-of-the-art methods for:

1) cervical cancer diagnosis by fusion of diagnosis from vertical segments of the epithelium to obtain an whole image-based diagnosis, and,

2) nuclei-based feature extraction and selection, ground truth study, and

3) multi-modal biomedical figure discrimination for application with content-based image retrieval.

To summarize, the research performed showed that fusion of data at feature- and decision-level is important for enhanced detection and classification. Feature-data fusion using feature groups including nuclei and nuclei related feature were used to obtain a better classification accuracy of histology image classification. Furthermore, the feature selection and extraction, at the feature-level provided a enhancement for feature groups to obtain a improved classification. Also, fusion with statistical methods were used for feature selection. Decision-level fusion is accomplished utilizing voting algorithm to obtain the final decision for whole epithelial image.

The research presented in this dissertation concludes that: 1) feature-level data fusion techniques are useful for dimension reduction, which in a result, can increase the prediction accuracy, 2) decision level data fusion techniques could be utilized for enhanced classification. 3) A refined feature extraction procedure can contribute in improving the

classification. Experimental results from this dissertation show that as compared to conventional image processing techniques, data fusion techniques at feature and decision level can provide more comprehensive image analysis and enhanced classification.

REFERENCES

[1] R. Faingold, A. Daneman, G. Tomlinson, P. S. Babyn, D. E. Manson, A. Mohanta, A. M. Moore, J. Hellmann, C. Smith, T. Gerstle, and J. H. Kim, "Necrotizing Enterocolitis: Assessment of Bowel Viability with Color Doppler US," *Radiology*, vol. 235, pp. 587–594, 5 2005.

[2] C. S. Restrepo, L. Diethelm, J. A. Lemos, E. Velásquez, T. A. Ovella, S. Martinez, J. Carrillo, and D. F. Lemos, "Cardiovascular Complications of Human Immunodeficiency Virus Infection," *RadioGraphics*, vol. 26, pp. 213–231, 1 2006.

[3] M. S. Taljanovic, T. B. Hunter, M. D. Miller, and J. E. Sheppard, "Gallery of Medical Devices," *RadioGraphics*, vol. 25, pp. 859–870, 5 2005.

[4] N. Hiller, S. Lieberman, T. Chajek-Shaul, J. Bar-Ziv, and D. Shaham, "Thoracic Manifestations of Behçet Disease at CT," *RadioGraphics*, vol. 24, pp. 801–808, 5 2004.

[5] H. P. Clark, W. F. Carson, P. V. Kavanagh, C. P. H. Ho, P. Shen, and R. J. Zagoria, "Staging and Current Treatment of Hepatocellular Carcinoma," *RadioGraphics*, vol. 25, pp. S3–S23, 10 2005.

[6] K. Awai, K. Murao, A. Ozawa, M. Komi, H. Hayakawa, S. Hori, and Y. Nishimura, "Pulmonary Nodules at Chest CT: Effect of Computer-aided Diagnosis on RadiologistsâĂŹ Detection Performance," *Radiology*, vol. 230, pp. 347–352, 2 2004.

[7] A. G. Blum, J.-P. Zabel, R. Kohlmann, T. Batch, K. Barbara, X. Zhu, G. Dautel, and F. Dap, "Pathologic Conditions of the Hypothenar Eminence: Evaluation with Multidetector CT and MR Imaging," *RadioGraphics*, vol. 26, pp. 1021–1044, 7 2006.

[8] H. Rhim, G. D. Dodd, K. N. Chintapalli, B. J. Wood, D. E. Dupuy, J. L. Hvizda, P. E. Sewell, and S. N. Goldberg, "Radiofrequency Thermal Ablation of Abdominal Tumors: Lessons Learned from Complications," *RadioGraphics*, vol. 24, pp. 41–52, 1 2004.

[9] P. J. Pickhardt, "Differential Diagnosis of Polypoid Lesions Seen at CT Colonography (Virtual Colonoscopy)," *RadioGraphics*, vol. 24, pp. 1535–1556, 11 2004.

[10] S. R. Prasad, H. Wang, H. Rosas, C. O. Menias, V. R. Narra, W. D. Middleton, and J. P. Heiken, "Fat-containing Lesions of the Liver: Radiologic-Pathologic Correlation," *RadioGraphics*, vol. 25, pp. 321–331, 3 2005.

[11] C. Bos, Y. Delmas, A. Desmoulière, A. Solanilla, O. Hauger, C. Grosset, I. Dubus, Z. Ivanovic, J. Rosenbaum, P. Charbord, C. Combe, J. W. M. Bulte, C. T. W. Moonen, J. Ripoche, and N. Grenier, "In Vivo MR Imaging of Intravascularly Injected Magnetically Labeled Mesenchymal Stem Cells in Rat Kidney and Liver," *Radiology*, vol. 233, pp. 781–789, 12 2004.

[12] T. Berrocal, M. Parrón, A. Álvarez-Luque, C. Prieto, and M. L. Santamaría, "Pediatric Liver Transplantation: A Pictorial Essay of Early and Late Complications," *RadioGraphics*, vol. 26, pp. 1187–1209, 7 2006.

[13] A. D. Levy, N. Patel, N. Dow, R. M. Abbott, M. Miettinen, and L. H. Sobin, "Abdominal Neoplasms in Patients with Neurofibromatosis Type 1: Radiologic-Pathologic Correlation," *RadioGraphics*, vol. 25, pp. 455–480, 3 2005.

[14] C. Wittram, M. M. Maher, A. J. Yoo, M. K. Kalra, J.-A. O. Shepard, and T. C. McLoud, "CT Angiography of Pulmonary Embolism: Diagnostic Criteria and Causes of Misdiagnosis," *RadioGraphics*, vol. 24, pp. 1219–1238, 9 2004.

[15] R. J. Stanley, S. De, D. Demner-Fushman, S. Antani, and G. R. Thoma, "An image feature-based approach to automatically find images for application to clinical decision support," *Computerized Medical Imaging and Graphics*, vol. 35, no. 5, pp. 365–372, 2011.

[16] B. Cheng, R. J. Stanley, S. De, S. Antani, and G. R. Thoma, "Automatic Detection of Arrow Annotation Overlays in Biomedical Images," *International Journal of Healthcare Information Systems and Informatics*, vol. 6, pp. 23–41, 34 2011.

[17] D. Hall and J. Llinas, "An introduction to multisensor data fusion," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 6–23, 1997.

[18] J. Manyika and H. F. Durrant-Whyte, *Data Fusion and Sensor Management: An Information-Theoretic Approach.* Ellis Horwood, 1994.

[19] D. L. D. L. Hall and S. A. H. McMullen, *Mathematical techniques in multi-sensor data fusion*. Artech House, 2004.

[20] D. Aneja and T. K. Rawat, "Fuzzy Clustering Algorithms for Effective Medical Image Segmentation," *International Journal of Intelligent Systems and Applications*, vol. 5, no. 11, pp. 55–61, 2013.

[21] T. Bayes, "An Essay Towards Solving a Problem in the Doctrines of Chances," *Philosophical Transactions*, vol. 53, no. 1764, pp. 370–418, 1763.

[22] I. J. Rudas, "Detection of abrupt changes. Theory and application," *Control Engineering Practice*, vol. 2, no. 4, pp. 729–730, 1994.

[23] E. Blasch, I. Kadar, J. Salerno, M. M. Kokar, S. Das, G. M. Powell, D. D. Corkill, and E. H. Ruspini, "Issues and challenges of knowledge representation and reasoning methods in situation assessment (Level 2 Fusion)," in *Defense and Security Symposium* (I. Kadar, ed.), vol. 6235, pp. 623510–623510, 5 2006.

[24] Z. Liu, D. S. Forsyth, J. P. Komorowski, K. Hanasaki, and T. Kirubarajan, "Survey: State of the art in NDE data fusion techniques," *IEEE Transactions on Instrumentation and Measurement*, vol. 56, pp. 2435–2451, 12 2007.

[25] V. Kumar, A. K. Abbas, N. Fausto, and J. C. Aster, *Robbins and Cotran Pathologic Basis of Disease*. Elsevier Health Sciences, 2014.

[26] J. Van Der Marel, W. G. V. Quint, M. Schiffman, M. M. van de Sandt, R. E. Zuna, S. T. Dunn, K. Smith, C. A. Mathews, M. A. Gold, J. Walker, and N. Wentzensen, "Molecular mapping of high-grade cervical intraepithelial neoplasia shows etiological dominance of HPV16," *International Journal of Cancer*, vol. 131, no. 6, pp. 946–53, 2012.

[27] L. He, L. R. Long, S. Antani, and G. Thoma, "Computer assisted diagnosis in histopathology," *Sequence and Genome Analysis: Methods and Applications*, vol. 3, pp. 271–287, 2010.

[28] Y. Wang, D. Crookes, O. S. Eldin, S. Wang, P. Hamilton, and J. Diamond, "Assisted diagnosis of cervical intraepithelial neoplasia (CIN)," *IEEE Journal on Selected Topics in Signal Processing*, vol. 3, pp. 112–121, 2 2009.

[29] S. M. Ismail, A. B. Colclough, J. S. Dinnen, D. Eakins, D. M. Evans, E. Gradwell, J. P. O'Sullivan, J. M. Summerell, and R. Newcombe, "Reporting cervical intra-epithelial neoplasia (CIN): intra- and interpathologist variation and factors associated with disagreement.," *Histopathology*, vol. 16, pp. 371–376, 4 1990.

[30] S. De, R. J. Stanley, C. Lu, R. Long, S. Antani, G. Thoma, and R. Zuna, "A fusion-based approach for uterine cervical cancer histology image classification," *Computerized Medical Imaging and Graphics*, vol. 37, no. 7-8, pp. 475–87, 2013.

[31] J. Eakins, J. Eakins, M. Graham, J. Eakins, M. Graham, and T. Franklin, "Content-based Image Retrieval," *LIBRARY AND INFORMATION BRIEFINGS*, vol. 85, pp. 1–15, 1999.

[32] K. Banerjee, "Uterine cervical cancer histology image feature extraction and classification," 2014.

[33] P. Guo, K. Banerjee, R. Stanley, R. Long, S. Antani, G. Thoma, R. Zuna, S. R. S. Frazier, R. H. Moss, and W. V. Stoecker, "Nuclei-Based Features for Uterine Cervical Cancer Histology Image Analysis with Fusion-based Classification," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, pp. 1595–1607, 10 2015.

[34] World Health Organization, "Comprehensive Cervical Cancer Control. A guide to essential practice," *WHO Library Cataloguing-in-Publication Data*, pp. 1–284, 2006.

[35] J. Jeronimo, M. Schiffman, R. L. Long, L. Neve, and S. Antani, "A tool for collection of region based data from uterine cervix images for correlation of visual and clinical variables related to cervical neoplasia," in *Proceedings. 17th IEEE Symposium on Computer-Based Medical Systems*, pp. 558–562, IEEE Comput. Soc, 2004.

[36] C. Molloy, C. Dunton, P. Edmonds, M. F. Cunnane, and T. Jenkins, "Evaluation of colposcopically directed cervical biopsies yielding a histologic diagnosis of CIN 1,2.," *Journal of lower genital tract disease*, vol. 6, no. 2, pp. 80–3, 2002.

[37] M. Guillaud, K. Adler-Storthz, A. Malpica, G. Staerkel, J. Matisic, D. Van Niekirk, D. Cox, N. Poulin, M. Follen, and C. MacAulay, "Subvisual chromatin changes in cervical epithelium measured by texture image analysis and correlated with HPV," *Gynecologic Oncology*, vol. 99, pp. 16–23, 12 2005.

[38] S. J. Keenan, J. Diamond, W. Glenn McCluggage, H. Bharucha, D. Thompson, P. H. Bartels, and P. W. Hamilton, "An automated machine vision system for the histological grading of cervical intraepithelial neoplasia (CIN)," *Journal of Pathology*, vol. 192, no. 3, pp. 351–62, 2000.

[39] C. Maurer and V. Raghavan, "A linear time algorithm for computing exact Euclidean distance transforms of binary images in arbitrary dimensions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 2, pp. 265–270, 2003.

[40] C. Rao, R. Toutenburg, H. Heumann, and S. Christian, *Linear Models and Generalizations Least Squares and Alternatives*. 2008.

[41] Mathworks, "Matlab (2015)."

[42] J. Borovicka, "Circle detection using Hough transforms Course Project: COMS30121- Image Processing and Computer Vision," tech. rep., 2003.

[43] R. Gonzalez and R. Woods, *Digital image processing*. 2002.

[44] F. P. Preparata and M. I. Shamos, *Computational geometry: an introduction*, vol. 47. 1985.

[45] C.-C. Chang and C.-j. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, pp. 1–39, 2011.

[46] W. J. Krzanowski, *Principles of Multivariate Analysis: A UserâĂŹs Perspective*. Oxford University Press, New York, Oxford, 1988.

[47] R.-E. Fan, P.-H. Chen, and C.-J. Lin, "Working Set Selection Using Second Order Information for Training Support Vector Machines," *J. Mach. Learn. Res.*, vol. 6, pp. 1889–1918, 12 2005.

[48] M. Pal, "Multinomial logistic regression-based feature selection for hyperspectral data," *International Journal of Applied Earth Observation and Geoinformation*, vol. 14, no. 1, pp. 214–220, 2012.

[49] T. Li, S. Zhu, and M. Ogihara, "Using discriminant analysis for multi-class classification: An experimental investigation," *Knowledge and Information Systems*, vol. 10, no. 4, pp. 453–472, 2006.

[50] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression*. No. 1, 2000.

[51] M. A. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *SIGKDD Explorations*, vol. 11, no. 1, pp. 10–18, 2009.

[52] A. Agresti, *An introduction to categorical Data Analysis*, vol. 22. 1996.

[53] M. Veta, P. J. van Diest, R. Kornegoor, A. Huisman, M. A. Viergever, and J. P. W. Pluim, "Automatic Nuclei Segmentation in H&E Stained Breast Cancer Histopathology Images," *PloS one*, vol. 8, no. 7, p. e70221, 2013.

[54] W. G. McCluggage, M. Y. Walsh, C. M. Thornton, P. W. Hamilton, A. Date, L. M. Caughley, and H. Bharucha, "Inter- and intra-observer variation in the histopathological reporting of cervical squamous intraepithelial lesions using a modified Bethesda grading system," *British journal of obstetrics and gynaecology*, vol. 105, pp. 206–10, 2 1998.

[55] G. S. Lodwick, "Diagnostic Signs and Minipatterns," *Proc Annu. Symp. Comput. Appl. Ded.Care*, vol. 3, pp. 1849–1850, 11 1980.

[56] J. R. Egner, "AJCC Cancer Staging Manual," *JAMA*, vol. 304, p. 1726, 10 2010.

[57] L. He, L. R. Long, S. Antani, and G. R. Thoma, "Histology image analysis for carcinoma detection and grading," *Computer Methods and Programs in Biomedicine*, vol. 107, no. 3, pp. 538–556, 2012.

[58] M. Mignotte, "Segmentation by fusion of histogram-based K-means clusters in different color spaces," *IEEE Transactions on Image Processing*, vol. 17, no. 5, pp. 780–787, 2008.

[59] P. Guo, *Cervical cancer histology image feature extraction and classification*. PhD thesis, Missouri University of Science and Technology, Rolla, Missouri, 2014.

[60] S. L. Cessie and J. C. V. Houwelingen, "Ridge Estimators in Logistic Regression," *Applied Statistics*, vol. 41, no. 1, p. 191, 1992.

[61] L. Breiman, "Manual on setting up, using, and understanding random forests v3. 1," *Technical Report, http://oz.berkeley.edu/users/breiman, Statistics Department University of California Berkeley, . . .* , p. 29, 2002.

[62] L. Breiman and A. Cutler, "Breiman and Cutler's random forests for classification and regression," *Package 'randomForest'*, p. 29, 2012.

[63] A. Agresti, *An Introduction to Categorical Data Analysis*. 2007.

[64] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression 2nd ed. New York John Willey & Sons*. New York: John Willey & Sons, 2000.

[65] S. De, R. J. Stanley, B. Cheng, S. Antani, R. Long, and G. Thoma, "Automated Text Detection and Recognition in Annotated Biomedical Publication Images," *Int. J. Healthc. Inf. Syst. Inform.*, vol. 9, pp. 34–63, 4 2014.

[66] B. Cheng, R. Wang, S. Antani, R. J. Stanley, and G. R. Thoma, "Graphical image classification combining an evolutionary algorithm and binary particle swarm optimization," p. 829703, 1 2012.

[67] D. A. Lindberg, B. L. Humphreys, and A. T. McCray, "The Unified Medical Language System.," *Methods Archive*, vol. 32, no. 4, pp. 281–291, 1993.

[68] M. Rahman, S. K. Antani, G. R. Thoma, and U. S. N. Library, "A Medical Image Retrieval Framework in Correlation Enhanced Visual Concept Feature Space," pp. 6–9, 2009.

[69] M. M. Rahman, S. K. Antani, and G. R. Thoma, "A classification-driven similarity matching framework for retrieval of biomedical images," in *MIR 2010 - Proceedings of the 2010 ACM SIGMM International Conference on Multimedia Information Retrieval*, pp. 147–154, 2010.

[70] Y. Xue, T. Cheng, X. Xu, Z. Gao, Q. Li, X. Liu, X. Wang, R. Song, X. Ju, and Q. Zhang, "High-accuracy and real-time 3D positioning, tracking system for medical imaging applications based on 3D digital image correlation," *Optics and Lasers in Engineering*, vol. 88, pp. 82–90, 2017.

[71] S. Zhang and D. Metaxas, "Large-Scale Medical Image Analytics: Recent Methodologies, Applications and Future Directions," *Medical Image Analysis*, vol. 33, pp. 98–101, 2016.

[72] M. Sundarapandian, R. Kalpathi, R. A. C. Siochi, and A. S. Kadam, "Lung diaphragm tracking in CBCT images using spatio-temporal MRF," *Computerized Medical Imaging and Graphics*, vol. 53, pp. 9–18, 2016.

[73] T. J. Bright, A. Wong, R. Dhurjati, E. Bristow, L. Bastian, R. R. Coeytaux, G. Samsa, V. Hasselblad, J. w. Williams, M. D. Musty, L. Wing, A. S. Kendrick, G. D. Sanders, and D. Lobach, "Annals of Internal Medicine Review Effect of Clinical Decision-Support Systems," *Annals of Internal Medicine*, vol. 157, no. 1, pp. 29–43, 2012.

[74] D. Zhang, A. Wong, M. Indrawan, and G. Lu, “Content-based Image Retrieval Using Gabor Texture Features,” *IEEE Transactions PAMI*, vol. 3656 LNCS, p. 13âĂŞ15, 2000.

[75] P. S. Hiremath and J. Pujari, “Content Based Image Retrieval based on Color , Texture and Shape features using Image and its complement,” *International Journal of Computer Science and Security*, vol. 1, no. 4, pp. 25–35, 2007.

[76] S. Ghosh and A. Ghosh, “Content Based Retrieval of Malaria Positive Images from a Clinical Database VIA Recognition in RGB Colour Space,” *Advances in Intelligent Systems and Computing*, vol. 248, 2014.

[77] P. H. Bugatti, D. S. Kaster, M. Ponciano-Silva, C. Traina, P. M. Azevedo-Marques, and A. J. M. Traina, “PRoSPer: Perceptual similarity queries in medical CBIR systems through user profiles,” *Computers in Biology and Medicine*, vol. 45, no. 1, pp. 8–19, 2014.

[78] G. W. Hruby, L. V. Rasmussen, D. Hanauer, V. L. Patel, J. J. Cimino, and C. Weng, “A multi-site cognitive task analysis for biomedical query mediation,” *International Journal of Medical Informatics*, vol. 93, pp. 74–84, 2016.

[79] S. Istephan and M. R. Siadat, “Unstructured medical image query using big data - An epilepsy case study,” *Journal of Biomedical Informatics*, vol. 59, pp. 218–226, 2016.

[80] E. Tramontana, “Graphic object feature extraction system based on cuckoo search algorithm,” 2016.

[81] M. Behrisch, B. Bach, M. Hund, M. Delz, L. von Ruden, J.-D. Fekete, and T. Schreck, “Magnostics: Image-based Search of Interesting Matrix Views for Guided Network Exploration,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 2626, no. c, pp. 1–1, 2016.

[82] F. Zhang, S. Member, Y. Song, and W. Cai, "Pairwise Latent Semantic Association for Similarity Computation in Medical Imaging," vol. 9294, no. c, pp. 1–12, 2015.

[83] H. Yuan and X.-p. Zhang, "Texture Image Retrieval Based on a Gaussian Mixture Model and Similarity Measure Using a Kullback Divergence," pp. 1867–1870, 2004.

[84] K. Jalaja, C. Bhagvati, B. L. Deekshatulu, and A. K. Pujari, "Texture Element Feature Characterizations for CBIR," vol. 500046, pp. 733–736, 2005.

[85] P. Mohapatra, S. Chakravarty, and P. K. Dash, "An improved cuckoo search based extreme learning machine for medical data classification," *Swarm and Evolutionary Computation*, vol. 24, pp. 25–49, 2015.

[86] H.-S. Oh and Y. Jung, "Cluster-based query expansion using external collections in medical information retrieval.," *Journal of biomedical informatics*, vol. 58, pp. 70–79, 2015.

[87] D. A. Szlosek and J. M. Ferretti, "Using Machine Learning and Natural Language Processing Algorithms to Automate the Evaluation of Clinical Decision Support in Electronic Medical Record Systems," no. August, 2016.

[88] M. Zalis and M. Harris, "Advanced Search of the Electronic Medical Record: Augmenting Safety and Efficiency in Radiology," *Journal of the American College of Radiology*, vol. 7, no. 8, pp. 625–633, 2010.

[89] L. Marco-Ruiz, C. Pedrinaci, J. Maldonado, L. Panziera, R. Chen, and J. G. Bellika, "Publication, discovery and interoperability of Clinical Decision Support Systems: A Linked Data approach," *Journal of Biomedical Informatics*, vol. 62, pp. 243–264, 2016.

[90] Y. Molenbruch, K. Braekers, A. Caris, and G. Vanden Berghe, "Multi-directional local search for a bi-objective dial-a-ride problem in patient transportation," *Computers & Operations Research*, vol. 77, pp. 58–71, 2017.

[91] A. M. Redd, A. V. Gundlapalli, G. Divita, M. E. Carter, L.-T. Tran, and M. H. Samore, "A Pilot Study of a Heuristic Algorithm for Novel Template Identification from {VA} Electronic Medical Record Text," *Journal of Biomedical Informatics*, pp. –, 2016.

[92] J. Jiang, J. Zheng, C. Zhao, J. Su, Y. Guan, and Q. Yu, "Clinical-decision support based on medical literature: A complex network approach," *Physica A: Statistical Mechanics and its Applications*, vol. 459, pp. 42–54, 2016.

[93] X. Liu, L. Huang, C. Deng, B. Lang, and D. Tao, "Query-Adaptive Hash Code Ranking for Large- Scale Multi-View Visual Search," vol. 25, no. 10, pp. 4514–4524, 2016.

[94] A. Mourão, F. Martins, and J. Magalhães, "Multimodal medical information retrieval with unsupervised rank fusion," *Computerized Medical Imaging and Graphics*, vol. 39, pp. 35–45, 2015.

[95] D. Demner-fushman, S. Antani, and G. R. Thoma, "Automatically Finding Images for Clinical Decision Support," pp. 139–144, 2007.

[96] J. P. Agrawal, B. J. Erickson, and C. E. Kahn, "Imaging Informatics: 25 Years of Progress," *IMIA Yearbook*, vol. 25, pp. S23–S31, 6 2016.

[97] C. Peterson, D. Rose, J. Mink, and D. Levitz, "Real-Time Monitoring and Evaluation of a Visual-Based Cervical Cancer Screening Program Using a Decision Support Job Aid," *Diagnostics*, vol. 6, p. 20, 5 2016.

[98] T. De Silva, S.-F. L. Lo, N. Aygun, D. M. Aghion, A. Boah, R. Petteys, A. Uneri, M. D. Ketcha, T. Yi, S. Vogt, G. Kleinszig, W. Wei, M. Weiten, X. Ye, A. Bydon, D. M. Sciubba, T. F. Witham, J.-P. Wolinsky, and J. H. Siewerdsen, "Utility of the LevelCheck Algorithm for Decision Support in Vertebral Localization," *SPINE*, vol. 41, pp. E1249–E1256, 10 2016.

[99] S. Antani, D. You, M. Simpson, M. Rahman, D. Demner-Fushman, and G. Thoma, "The Role of Image Modality and Visual Characteristics in Archiving Biomedical Images," in *Proc of IS&T Archiving Conference.*, Society for Imaging Science and Technology, 2013.

[100] M. Rahman, D. You, M. S. Simpson, S. K. Antani, D. Demner-Fushman, and G. R. Thoma, "An Interactive Image Retrieval Framework for Biomedical Articles Based on Visual Region-of- Interest (ROI) Identification and Classification," in *2012 IEEE Second International Conference on Healthcare Informatics, Imaging and Systems Biology*, pp. 50–50, IEEE, 9 2012.

[101] M. S. Simpson, D. You, M. M. Rahman, Z. Xue, D. Demner-Fushman, S. Antani, and G. Thoma, "Literature-based biomedical image classification and retrieval," *Computerized Medical Imaging and Graphics*, vol. 39, pp. 3–13, 1 2015.

[102] J. Kalpathy-Cramer, A. G. S. de Herrera, D. Demner-Fushman, S. Antani, S. Bedrick, and H. Müller, "Evaluating performance of biomedical image retrieval systems–an overview of the medical image retrieval task at ImageCLEF 2004-2013.," *Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society*, vol. 39, pp. 55–61, 1 2015.

[103] S. Vajda, D. You, S. K. Antani, and G. R. Thoma, "Label the many with a few: Semi-automatic medical image modality discovery in a large image collection," in *2014 IEEE Symposium on Computational Intelligence in Healthcare and e-health (CICARE)*, pp. 167–173, IEEE, 12 2014.

[104] Z. Xue, M. M. Rahman, S. Antani, L. R. Long, D. Demner-Fushman, and G. R. Thoma, "Modality Classification for Searching Figures in Biomedical Literature," in *2016 IEEE 29th International Symposium on Computer-Based Medical Systems (CBMS)*, pp. 152–157, IEEE, 6 2016.

[105] Z. Xue, D. You, S. Chachra, S. Antani, L. R. Long, D. Demner-Fushman, and G. R. Thoma, "Extraction of endoscopic images for biomedical figure classification [9418-23]," in *Proc. SPIE. 9418, Medical Imaging 2015: PACS and Imaging Informatics: Next Generation and Innovations, 94180P*, 2015.

[106] Z. Xue, D. You, S. Antani, L. R. Long, D. Demner-Fushman, and G. R. Thoma, "Classification of visual signs in abdominal CT image figures in biomedical literature," vol. 9039, p. 90390L, International Society for Optics and Photonics, 3 2014.

[107] Z. Xue, D. You, S. Candemir, S. Jaeger, S. Antani, L. R. Long, and G. R. Thoma, "Chest X-ray Image View Classification," in *2015 IEEE 28th International Symposium on Computer-Based Medical Systems*, pp. 66–71, IEEE, 6 2015.

[108] Z. Xue, S. Antani, L. R. Long, D. Demner-Fushman, and G. R. Thoma, "Body Segment Classification for Visible Human Cross Section Slices," in *2014 IEEE 27th International Symposium on Computer-Based Medical Systems*, pp. 199–204, IEEE, 5 2014.

[109] D. You, M. Simpson, S. Antani, D. Demner-Fushman, and G. R. Thoma, "A robust pointer segmentation in biomedical images toward building a visual ontology for biomedical article retrieval," vol. 8658, p. 86580Q, International Society for Optics and Photonics, 2 2013.

[110] D. You, M. M. Rahman, S. Antani, D. Demner-Fushman, and G. R. Thoma, "Text- and content-based biomedical image modality classification," vol. 8674, p. 86740L, International Society for Optics and Photonics, 3 2013.

[111] D. You, M. Simpson, S. Antani, D. Demner-Fushman, and G. R. Thoma, "Annotating image ROIs with text descriptions for multimodal biomedical document retrieval," vol. 8658, p. 86580D, International Society for Optics and Photonics, 2 2013.

[112] D. You, S. Antani, D. Demner-Fushman, and G. R. Thoma, "Does Figure-Text Improve Biomedical Article Retrieval? A Pilot Study," in *2014 IEEE 27th International Symposium on Computer-Based Medical Systems*, pp. 471–472, IEEE, 5 2014.

[113] "No Title."

[114] F. Nourbakhsh, P. B. Pati, and A. G. Ramakrishnan, "Text Localization and Extraction from Complex Gray Images," pp. 776–785, Springer, Berlin, Heidelberg, 2006.

[115] H. A. Jalab, "Image retrieval system based on color layout descriptor and Gabor filters," in *2011 IEEE Conference on Open Systems*, pp. 32–36, IEEE, 9 2011.

[116] A. Oliva and A. Torralba, "Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope," *International Journal of Computer Vision*, vol. 42, no. 3, pp. 145–175, 2001.

VITA

Peng Guo was born in city of Jilin in the province of Jilin, China in 1989. He did his schooling at the Tianjin Road Elementary School (1996-2002), 7th Middle School (2002-2005), in Jilin, and No.1 Middle School (2005-2008) before going to Northeast Electrical University in his hometown, for his Bachelor of Science degree in Electrical Engineering from the Department of Automation (2012). In July of the same year, he entered Missouri University of Science and Technology, where he completed his MS in Electrical Engineering (2014) and continued his PhD in the same program. He worked as an intern with research group of Lister Hill National Center for Biomedical Communications at National Institutes of Health (NIH) during the summer of 2015, 2016 and 2017. He received his PhD degree in Electrical Engineering from Missouri University of Science and Technology in May 2018.