

---

Doctoral Dissertations

Student Theses and Dissertations

---

Fall 2020

## Deep learning for digitized histology image analysis

Sudhir Sornapudi

Follow this and additional works at: [https://scholarsmine.mst.edu/doctoral\\_dissertations](https://scholarsmine.mst.edu/doctoral_dissertations)



Part of the [Artificial Intelligence and Robotics Commons](#), [Bioimaging and Biomedical Optics Commons](#), and the [Computer Engineering Commons](#)

**Department: Electrical and Computer Engineering**

---

### Recommended Citation

Sornapudi, Sudhir, "Deep learning for digitized histology image analysis" (2020). *Doctoral Dissertations*. 3110.

[https://scholarsmine.mst.edu/doctoral\\_dissertations/3110](https://scholarsmine.mst.edu/doctoral_dissertations/3110)

This thesis is brought to you by Scholars' Mine, a service of the Missouri S&T Library and Learning Resources. This work is protected by U. S. Copyright Law. Unauthorized use including reproduction for redistribution requires the permission of the copyright holder. For more information, please contact [scholarsmine@mst.edu](mailto:scholarsmine@mst.edu).

DEEP LEARNING FOR DIGITIZED HISTOLOGY IMAGE ANALYSIS

by

SUDHIR SORNAPUDI

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

COMPUTER ENGINEERING

2020

Approved by:

R. Joe Stanley, Advisor

William V. Stoecker

Kurt Kosbar

Maciej J. Zawodniok

V. A. Samaranayake

© 2020

Sudhir Sornapudi

All Rights Reserved

## **PUBLICATION DISSERTATION OPTION**

This dissertation consists of the following four articles:

Paper I, found on pages 7–36, has been published in the *Journal of Pathology Informatics*.

Paper II, found on pages 37–64, has been published in the *Journal of Pathology Informatics*.

Paper III, found on pages 65–95, has been submitted to the *Journal of Pathology Informatics*.

Paper IV, found on pages 96–120, has been submitted to the *Journal of Pathology Informatics*.

## ABSTRACT

Cervical cancer is the fourth most frequent cancer that affects women worldwide. Assessment of cervical intraepithelial neoplasia (CIN) through histopathology remains as the standard for absolute determination of cancer. The examination of tissue samples under a microscope requires considerable time and effort from expert pathologists. There is a need to design an automated tool to assist pathologists for digitized histology slide analysis. Pre-cervical cancer is generally determined by examining the CIN which is the growth of atypical cells from the basement membrane (bottom) to the top of the epithelium. It has four grades, including: Normal, CIN1, CIN2, and CIN3. In this research, different facets of an automated digitized histology epithelium assessment pipeline have been explored to mimic the pathologist diagnostic approach. The entire pipeline from slide to epithelium CIN grade has been designed and developed using deep learning models and imaging techniques to analyze the whole slide image (WSI). The process is as follows: 1) identification of epithelium by filtering the regions extracted from a low-resolution image with a binary classifier network; 2) epithelium segmentation; 3) deep regression for pixel-wise segmentation of epithelium by patch-based image analysis; 4) attention-based CIN classification with localized sequential feature modeling. Deep learning-based nuclei detection by superpixels was performed as an extension of our research. Results from this research indicate an improved performance of CIN assessment over state-of-the-art methods for nuclei segmentation, epithelium segmentation, and CIN classification, as well as the development of a prototype WSI-level tool.

## ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisor, Dr. R. Joe Stanley, for the continual support, guidance, and encouragement he has provided throughout the journey of my Ph.D. study. He always encouraged me to take up various internships, which have helped me grow professionally as a researcher. His guidance helped me during the research and writing of this dissertation. Besides my advisor, I am also grateful to Dr. William V. Stoecker, Dr. Sameer Antani, Mr. Rodney Long, Dr. Kurt Kosbar, Dr. Maciej J. Zawoniok, and Dr. V. A. Samaranayake not only for their insightful comments and encouragement, but also for their hard questions, which motivated me to widen my research from various perspectives.

I would like to thank my life partner, Dr. P. Deepika Devada, for her relentless encouragement and moral support throughout my Ph.D. Thanks to all my colleagues: Jason H., Haider A., Norsang L., and Anand N., for their inspiring thoughts and company during my work at Missouri University of Science and Technology. Their interactions and shared ideas helped me work on various approaches to a problem.

I am grateful to my parents, whose love and affection are always a blessing. It was their encouragement and support that constantly gave me the courage to lead my life.

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

## TABLE OF CONTENTS

	Page
PUBLICATION DISSERTATION OPTION .....	iii
ABSTRACT .....	iv
ACKNOWLEDGMENTS .....	v
LIST OF ILLUSTRATIONS .....	xi
LIST OF TABLES .....	xiv
NOMENCLATURE .....	xvi
 SECTION	
1. INTRODUCTION .....	1
1.1. OVERVIEW .....	2
1.2. PROBLEM DESCRIPTION .....	2
1.3. SUMMARY OF CONTRIBUTIONS .....	4
1.3.1. Deep Learning Nuclei Detection in Digitized Histology Images by Superpixels.....	4
1.3.2. EpithNet: Deep Regression for Epithelium Segmentation in Cervical Histology Images. ....	4
1.3.3. DeepCIN: Attention-based Cervical Histology Image Classification with Sequential Feature Modelling for Pathologist-Level Accuracy. ....	5
1.3.4. Fully Automated End-to-end Cervical Histology Whole Slide Image Diagnosis Toolbox.....	6
 PAPER	
I. DEEP LEARNING NUCLEI DETECTION IN DIGITIZED HISTOLOGY IMAGES BY SUPERPIXELS .....	7
ABSTRACT .....	7

1. INTRODUCTION.....	8
2. METHODS.....	13
2.1. PRE-PROCESSING .....	14
2.2. SUPERPIXEL EXTRACTION .....	16
2.3. DATA GENERATION.....	17
2.4. CONVOLUTIONAL NEURAL NETWORK.....	19
3. EXPERIMENTAL RESULTS AND ANALYSIS .....	24
3.1. EXPERIMENTAL RESULTS .....	24
3.2. ANALYSIS OF RESULTS .....	25
3.3. COMPARISON OF RESULTS.....	27
4. CONCLUSION .....	31
ACKNOWLEDGEMENTS .....	31
REFERENCES.....	32
II. EPITHNET: DEEP REGRESSION FOR EPITHELIUM SEGMENTATION IN CERVICAL HISTOLOGY IMAGES.....	37
ABSTRACT .....	37
1. INTRODUCTION.....	38
2. METHODS.....	42
2.1. DATA .....	43
2.2. OVERVIEW OF PROPOSED SEGMENTATION METHOD .....	43
2.3. INPUT IMAGE DATA .....	44
2.4. REGRESSION MODEL.....	45
2.5. POST-PROCESSING.....	49
3. EXPERIMENTS .....	52

3.1. EXPERIMENTAL MODELS .....	52
3.2. EXPERIMENTAL RESULTS .....	55
4. DISCUSSION .....	57
5. CONCLUSION .....	60
ACKNOWLEDGEMENTS .....	62
REFERENCES.....	62
III. DEEPCIN: ATTENTION-BASED CERVICAL HISTOLOGY IMAGE CLASSIFICATION WITH SEQUENTIAL FEATURE MODELLING FOR PATHOLOGIST-LEVEL ACCURACY .....	65
ABSTRACT .....	65
1. INTRODUCTION.....	66
2. METHODOLOGY .....	70
2.1. LOCALIZATION.....	71
2.2. SEGMENT-LEVEL SEQUENCE GENERATION.....	73
2.3. IMAGE-LEVEL CLASSIFICATION.....	76
2.4. TRAINING .....	78
3. EXPERIMENTS .....	79
3.1. DATASET AND EVALUATION METRICS .....	79
3.2. IMPLEMENTATION DETAILS .....	81
3.3. ABLATION STUDIES .....	82
4. PERFORMANCE OF DEEPCIN .....	86
5. DISCUSSION .....	89
6. CONCLUSION .....	90
ACKNOWLEDGMENTS.....	92

REFERENCES.....	92
IV. AUTOMATED CERVICAL DIGITIZED HISTOLOGY WHOLE SLIDE IMAGE ANALYSIS TOOLBOX.....	96
ABSTRACT.....	96
1. INTRODUCTION.....	97
2. METHODOLOGY.....	101
2.1. EPITHELIUM DETECTION.....	101
2.1.1. ROI Extraction.....	102
2.1.2. Epithelium Detection Network.....	103
2.2. EPITHELIUM SEGMENTATION.....	105
2.3. LOCALIZATION.....	106
2.4. CIN CLASSIFICATION.....	107
3. EXPERIMENTS.....	107
3.1. DATA.....	107
3.2. IMPLEMENTATION DETAILS.....	109
3.3. EVALUATION METRICS.....	110
4. RESULTS.....	111
4.1. PERFORMANCE OF EPITHELIUM DETECTION NETWORK.....	111
4.2. PERFORMANCE OF CIN CLASSIFICATION MODEL.....	113
5. DISCUSSION.....	115
6. CONCLUSION.....	117
ACKNOWLEDGMENTS.....	118
REFERENCES.....	118

## SECTION

2. SUMMARY AND CONCLUSIONS.....	121
BIBLIOGRAPHY.....	122
VITA.....	125

## LIST OF ILLUSTRATIONS

PAPER I	Page
Figure 1. CIN grades left to right: Normal, CIN1, CIN 2, CIN 3.....	9
Figure 2. Proposed methodology.....	13
Figure 3. A portion of original image with superpixels.....	15
Figure 4. Generation of 16x16x3 RGB image from superpixel.....	16
Figure 5. Images with lighter nuclei (left), darker nuclei with lighter cytoplasm (center), darker nuclei with thicker cytoplasm (right). .....	18
Figure 6. Generation of training dataset (left) and test dataset (right).....	18
Figure 7. Sample 16 x 16 x 3 RGB images and their 16 x 16 V-plane images. ....	20
Figure 8. CNN Architecture.....	21
Figure 9. Training loss, validation accuracy and validation loss vs. epochs. ....	22
Figure 10. 32x3x3 CNN filters and 32x14x14 convolved output in first layer. ....	23
Figure 11. Nuclei masks (green) superimposed on the original image.....	25
Figure 12. Examples of false positive (left) and false negative (right) results. ....	26
Figure 13. Nuclei detected even under red stains. ....	27
<b>PAPER II</b>	
Figure 1. (a) DM image at 10X magnification with corresponding (b) manually generated mask.....	39
Figure 2. Epithelium analysis process used in previous research based on a manually segmented epithelium. ....	40
Figure 3. Generation of labels.....	45
Figure 4. EpithNet architecture.....	47
Figure 5. Original image split into tiles. ....	48

Figure 6. Predicted mask from each tile of original image.....	49
Figure 7. Post-processing: (a) clean mask and (b) mask edge smoothing.....	50
Figure 8. Segmentation contour.....	51
Figure 9. EpithNet-mc architecture.....	54
Figure 10. Boxplot of EpithNet-mc model with distribution of the metrics on 311 images. ....	58
Figure 11. Segmentation results.....	59

### PAPER III

Figure 1. Sections of epithelium region with increasing CIN severity (from (b)-(d)) showing delayed maturation with increase in immature atypical cells from bottom-to-top. ....	67
Figure 2. Overview of DeepCIN model.....	71
Figure 3. Localized vertical segment generation from an epithelial image.....	72
Figure 4. Segment-level sequence generator network with two-stage encoder structures. ....	74
Figure 5. Attention-based fusion network for epithelial image-level classification.....	77
Figure 6. Results of DeepCIN.....	87
Figure 7. Sankey diagram – based on the combined test results from the 5-fold cross-validation.....	88

### PAPER IV

Figure 1. Graphical overview of the proposed toolbox. ....	98
Figure 2. Overview of the proposed toolbox .....	100
Figure 3. Steps for ROI extraction.....	102
Figure 4. Mapping of high-resolution ROI (right) to its low-resolution image (left).....	103
Figure 5. Filtering of epithelium ROIs with the results from the epithelium detection network. ....	104

Figure 6. (a) Epithelium segmentation mask overlaid as a contour on the epithelium ROI. (b) Vertical segments generation through the localization process. .... 106

Figure 7. Examples of epithelium detection results..... 112

## LIST OF TABLES

PAPER I	Page
Table 1. Nuclei detection results using the deep learning superpixel approach. <sup>1</sup> .....	28
Table 2. Cervical nuclear detection vs. Current deep learning superpixel approach. <sup>1</sup> .....	28
Table 3. General nuclei detection results vs. Current deep learning superpixel approach. <sup>1</sup> .....	29
 PAPER II	
Table 1. Complexity of baseline, UNet-64 and the proposed models. ....	55
Table 2. Results on 311 cervical histology test data. ....	56
 PAPER III	
Table 1. Segment-level sequence generator model architecture.....	75
Table 2. Class label distribution from 453 epithelial images.....	80
Table 3. Ablation study on segment widths.....	83
Table 4. Ablation study on stage I encoder models. ....	84
Table 5. Ablation study on stage II encoder models.....	85
Table 6. Ablation study on fusion techniques.....	85
Table 7. Comparison with state-of-the-art models. ....	86
Table 8. 5-fold cross-validation results with different scoring schemes. ....	89
Table 9. CIN classification results on 224 image-set. ....	89
 PAPER IV	
Table 1. Epithelium detection network architecture.....	105
Table 2. Data distribution for epithelium detection.....	108
Table 3. Subset of epithelium ROI images for evaluating CIN classification.....	109

Table 4. Epithelium detection results.....	112
Table 5. CIN classification results on OU15-set. ....	113
Table 6. CIN classification results on OU16-set. ....	113
Table 7. CIN classification results on the combined set.....	115
Table 8. Benchmark CIN classification results [18]. ....	115

## NOMENCLATURE

Symbol	Description
$r_{min}$	Minimum pixel intensity
$r_{max}$	Maximum pixel intensity
$L_{\mu}$	Categorical accuracy
$t_{i,c}$	Target class
$p_{i,c}$	Prediction class
$TP$	True positive
$FP$	False positive
$TN$	True negative
$FN$	False negative
$\mu$	Accuracy
$DSC$	Dice coefficient
$F1$	F1 score
$JAC$	Jaccard index
$P$	Number of patches
$n$	number of rows in patch image
$m$	number of columns in patch image
$s$	stride
$p_{mask}$	Patch mask
$p$	number of rows in tile image
$q$	number of columns in tile image

$M$	number of rows in the original image
$N$	number of columns in the original image
$n_t$	number of tiles
$n_r$	rows in tile grid
$n_c$	columns in tile grid
$r$	number of points
$B$	Bezier curve
$J$	Jaccard index
$PA$	Pixel accuracy
$MI$	Mean intersection over union (IOU)
$FWI$	Frequency weighted IOU
$\kappa$	Cohen's kappa value
$I_{epth}$	Epithelial image
$I_{vsi}$	$i^{\text{th}}$ vertical segment
$vs_i$	$i^{\text{th}}$ logit sequence vector
$\theta$	Parameters of a model
$h_i$	hidden state
$\alpha_i$	Attentional weight of $i^{\text{th}}$ vertical segment
$v_I$	Image feature vector
$p_I$	Image class probability
$L_k$	Training loss of label $k$
$q_k$	weight of label $k$
$P$	Precision

$R$	Recall
$ACC$	Classification accuracy
$AUC$	Area under ROC
$AP$	Average Precision
$MCC$	Matthews correlation coefficient
$Sp$	Specificity
$Se$	Sensitivity
$H_{mean}$	Harmonic mean

## 1. INTRODUCTION

Cervical cancer is a cancer that occurs in the cells of the female cervix. It is caused by various strains of sexually transmitted infection, called human papillomavirus (HPV) [1]. The virus attacks the cervical cells and mutates some cells, which then become cancerous cells. According to the American Cancer Society, it is estimated that within the United States in 2020 about 13,800 women will be diagnosed with cervical cancer and about 4,290 are going to die from the disease [2]. Globally, around 570,000 new cases were developed in 2018, and approximately 90% of these deaths were recorded in low- and middle-income countries [3] due to lower availability of expert doctors and awareness for early screening. The WHO director general made a global call for the elimination of cervical cancer as a public health priority [4].

The risk of cancer can be reduced by having screening tests like the Pap test [5]. If abnormal cells are observed in the Pap test, the doctor may cut a sample of tissue from the cervix (biopsy), and a pathologist will examine it under a microscope (histopathology) [6]. The pathologist looks for CIN, cervical intraepithelial neoplasia (also called, cervical dysplasia), which is the growth of precancerous cervical cells in the epithelium that can potentially lead to cervical cancer. CIN is usually graded as CIN 1 (mild epithelial dysplasia), CIN 2 (moderate dysplasia), and CIN 3 (severe dysplasia). CIN 3 is carcinoma in-situ, spreading the entire thickness of the epithelium. The absence of CIN is considered normal [7]. This process is subject to inter- and intra-observer variability in the grading of cervical lesions [8].

## 1.1. OVERVIEW

Cervical cancer is one of the deadliest cancers that affects women worldwide. Cervical histopathology is one of the methods used by pathologists cancer diagnosis. The examination is performed at the tissue level with a light microscope to determine the severity of the cancer. A deeper analysis of cervical whole slide image (WSI) analysis is proposed, using novel techniques such as detecting cell nuclei, identifying and segmenting epithelium, and classifying cervical pre-cancerous stage. All of these proposed techniques are aggregated to design a toolbox that can use deep learning methods to analyze the digitized glass slide (WSI) and provide useful information about the cancerous regions and automatically diagnose cervical cancer. This tool can help assist the pathologists, and it can also help in preventing cancer through early detection.

## 1.2. PROBLEM DESCRIPTION

The advent of whole slide digital image scanners has paved a way for digital pathology that resulted in a substantial amount of clinical and research interests [9]. Effective computer vision approaches are proposed to enhance the quantitative and objective level of pathological research. Cervical cancer is a malignant uterine tumor that needs histopathological diagnosis and a detailed examination of cervical tissue and cells. The cervical lesions occur on the squamous epithelium, and hence there is a need for extensive epithelial analysis. The analysis can be performed through the detection and segmentation of the epithelium and its cells, which can help classify the cervical precancerous lesions through the estimation of CIN grades. The feature extraction methods provide crucial details for the epithelium analysis. These mainly include texture-

based features [10][11][12], shape-based features [13][12][11], color-based features [10][13][14] and deep learning features [13][14][15][16][17][18]. The segmentation methods include techniques based on threshold [19][20], edges [20][21], regions [22][23] and machine learning with support vector machine (SVM) [12], k-means clustering [13], and convolution neural networks (CNN) [25][26]. The CIN classification approaches were studied through traditional machine learning methods like fusion-based SVM and linear discriminant analysis (LDA) [27], enhancements features on SVM and LDA [28], LDA [29], and statistical approach [30], as well as deep learning methods with multi-layer hidden conditional random fields [10], ensembled transfer learning [16], spatial pattern correlation [25], and attention-based sequential feature modeling [31]. A detailed review of cervical image analysis is presented by Li *et. al.* [32].

In this dissertation, deep learning methods for automated cervical histology image analysis and diagnosis are investigated. Initially, superpixel-based nuclei identification in the epithelium regions is studied, since nuclei carry crucial information for CIN. Regression-based pixel-wise segmentation of the epithelium is proposed to separate the epithelium from the unwanted background. The epithelium is locally analyzed for an attention-based classification of cervical cancer for automated diagnosis. Finally, with a newly proposed epithelium detection model and with the help of the above-mentioned studies, an end-to-end prototype toolbox is designed to classify the severity of cervical cancer from a raw WSI.

### 1.3. SUMMARY OF CONTRIBUTIONS

This dissertation is comprised of four journal papers as listed in the publications list. The unique contributions from each article can be summarized as follows:

**1.3.1. Deep Learning Nuclei Detection in Digitized Histology Images by Superpixels.** Epithelial cell nuclei analysis is an initial critical step for cervical microscopy image analysis. In this study, superpixels are used as a base to extract localized image data. The superpixels are created using a K-means clustering-based simple linear iterative clustering (SLIC) algorithm [33] that is relatively faster and considers color and spatial proximity. V-plane (value plane from HSV color plane) is identified as the best image descriptor for identifying the nuclei. A custom shallow convolution neural network (CNN) is designed to model superpixel-level nuclei classifier. The CNN is a patch-based binary classifier that aims at categorizing the superpixel patches as nuclei and non-nuclei regions. The resulting information is consolidated through labeling the superpixels that would create a nuclei mask highlighting the nuclei regions. The object-based nuclei detection accuracy of 95.97% is achieved, which surpasses the state-of-the-art methods.

**1.3.2. EpithNet: Deep Regression for Epithelium Segmentation in Cervical Histology Images.** The cervical epithelium is the region of interest for pathologists, as it contains abnormally growing cervical cancer cells. There is a need for automated epithelium segmentation to facilitate the CIN diagnosis. In this paper, constructing a small and efficient CNN model for epithelium segmentation, named EpithNet, was explored. Patch-based regression CNN models are proposed and studied at three different spatial image dimensions. This is a pixel-level probability estimator network that aims to

estimate how probable it is that a given pixel is an epithelium pixel by analyzing a patch image surrounding that pixel. It is observed that the technique maximizes region-based information to improve pixel-wise probability estimates. Additionally, using a memory-optimized workflow testing at the pixel-level and an approximated Bezier curve to generate a smooth contour for the epithelium segmentation mask was explored. A concatenation of the three EpithNet models is performed, and the model resulted in providing the best results with a 94.0% Jaccard index compared to the benchmark model.

**1.3.3. DeepCIN: Attention-based Cervical Histology Image Classification with Sequential Feature Modelling for Pathologist-Level Accuracy.** Epithelial image analysis is extremely helpful to a pathologist when it can automatically classify the CIN grade. In this study, the DeepCIN network pipeline was proposed to locally analyze precancer CIN progression from the basement membrane to the top of the epithelium and model a CIN classifier. The model employs a novel methodology with a hierarchical image from localized regions to the entire epithelium image. The localized images are created as a series of standard width vertical segments with reference to a medial axis. The pipeline is comprised of a two-fold learning process. A weakly-supervised segment-level sequence generator is built as an image-to-sequence two-stage model. The sequences are consolidated and fused with a many-to-one attentional recurrent neural network (RNN) for whole epithelium image CIN classification. Additionally, the contributions by each vertical segment are identified to provide an in-depth analysis of the abnormality in the high-resolution epithelium image. Performance accuracy of 88.5% with accuracy under the ROC curve of 96.5% and Cohen's kappa score of 81.5% was

achieved by the proposed DeepCIN pipeline overperforming the state-of-the-art models with a pathologist-level accuracy.

**1.3.4. Fully Automated End-to-end Cervical Histology Whole Slide Image Diagnosis Toolbox.** The examination of biopsied tissue specimens under a microscope requires an expert pathologist to determine the CIN grade. This examination procedure needs a lot of time and effort. In this study, a fully automated end-to-end CIN diagnosis prototype toolbox was developed. The toolbox can directly scan a cervical whole slide image (WSI) and process it in four intermediate steps. First, the WSI is processed at its low-resolution to extract the outer region of interests (ROIs). These ROIs coordinates were recorded to extract the high-resolution image blocks. A CNN-based epithelium detection network was designed to filter and identify the epithelium-containing ROIs. Second, the epithelium mask is created from the segmentation of the epithelial ROIs using the EpiNet-64 model presented in our previous work [26]. Third, localized standard width vertical segments were generated about the medial axis drawn with the help of the segmented epithelial ROIs [31]. Fourth, attention-based sequence feature modeling (DeepCIN) [31] is performed to automatically classify the CIN grade for each epithelium ROI. The toolbox is advantageous in providing useful intermediate results, CIN classification for portions of the epithelium, and the CIN grade for the entire WSI with a voting mechanism. The results indicate that the toolbox can be easily scaled up for the real-world clinical setting and help to assist the pathologist expert in identifying the cells.

## PAPER

### I. DEEP LEARNING NUCLEI DETECTION IN DIGITIZED HISTOLOGY IMAGES BY SUPERPIXELS

Sudhir Sornapudi<sup>1</sup>, R. Joe Stanley<sup>1</sup>, William V. Stoecker<sup>5</sup>, Haidar Almubarak<sup>1</sup>, Rodney Long<sup>2</sup>, Sameer Antani<sup>2</sup>, George Thoma<sup>2</sup>, Rosemary Zuna<sup>3</sup>, Shelliane R. Frazier<sup>4</sup>

<sup>1</sup>Department of Electrical and Computer Engineering,  
Missouri University of Science and Technology, Rolla, MO

<sup>2</sup>Lister Hill National Center for Biomedical Communications,  
National Library of Medicine, National Institutes of Health, DHHS, Bethesda, MD

<sup>3</sup>Department of Pathology,  
University of Oklahoma Health Sciences Center, Oklahoma City, OK

<sup>4</sup>Surgical Pathology Department,  
University of Missouri Hospitals and Clinics, Columbia, MO

<sup>5</sup>Stoecker & Associates, Rolla, MO

### ABSTRACT

Advances in image analysis and computational techniques have facilitated the automatic detection of critical features in histopathology images. Detection of nuclei is critical for squamous epithelium cervical intraepithelial neoplasia (CIN) classification into Normal, CIN1, CIN2, and CIN3 grades. In this study, a deep learning-based nuclei segmentation approach is investigated based on gathering localized information through the generation of superpixels using a simple linear iterative clustering (SLIC) algorithm and training with a convolutional neural network. The proposed approach was evaluated on a data set of 133 digitized histology images and achieved an overall nuclei detection (object-based) accuracy of 95.97%, with demonstrated improvement over imaging- and clustering-based benchmark techniques.

*Index Terms*—Cervical cancer, cervical intraepithelial neoplasia, segmentation, deep learning, convolutional neural network, superpixels, image processing.

## 1. INTRODUCTION

The reconstruction of medical images into digital form has propelled the fields of medical research and clinical practice [1]. Image processing for histopathology image applications still has numerous challenges to overcome, especially in accurate nuclei detection.

Cervical cancer is the fourth most prevalent female cancer globally [2]. Over 500,000 new cases of this cancer are reported annually, especially in Africa; over half of this total eventuates in death [2]. There is a cure for cervical cancer if it is detected early. The gold standard for early cervical cancer diagnosis is the microscopic evaluation of histopathology images by a qualified pathologist [3][4][5][6]. The severity of cervical cancer increases as the immature atypical cells in the epithelium region increase. Based on this observation, the cancer affecting squamous epithelium is classified as normal or one of three increasingly pre-malignant grades of cervical intraepithelial neoplasia (CIN): CIN1, CIN2, and CIN3 [4][5][6] (Figure 1). Normal means there is no CIN; CIN1 corresponds to mild dysplasia (abnormal change); CIN2 denotes moderate dysplasia; and CIN3 corresponds to severe dysplasia.

With increasing CIN grade, the epithelium has been observed to show delayed maturation with an increase in immature atypical cells from bottom (basal membrane) to top of the epithelium region [7][8][9][10][11]. This can be observed from Figure 1.

Atypical immature cells are most dense in the bottom region of the epithelium for CIN1(Figure 1(b)). For CIN2, two-thirds of the bottom region is affected by the atypical immature cells (Figure 1(c)). Finally, for CIN3, the atypical immature cells are densely spread over the whole epithelium region (Figure 1(d)).

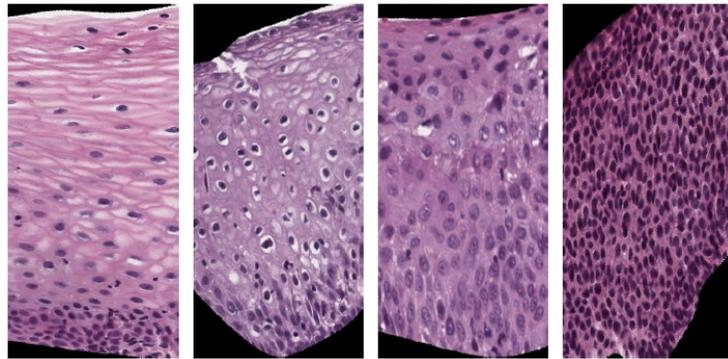


Figure 1. CIN grades left to right: Normal, CIN1, CIN 2, CIN 3.

At present, cervical tissue is analyzed manually by pathologists with significant experience with cervical cancer. These pathology specialists are few; and it takes considerable time to scan the tissue slides. This calls for automatic histology image classification, which could alleviate scarce professional resources for image classification, particularly in developing countries where the burden of cervical cancer is the greatest. A critical challenge for automatic classification is the accurate identification of nuclei, the small dark structures which undergo changes as the CIN progresses (Figure 1).

Epithelial nuclei provide critical features needed to classify cervical images. Although CIN grade classification can be done by applying deep learning techniques directly on the image data without the use of nuclei-based features, the accuracy of the

classification can be further improved by fusing a feature based trained neural network models with the deep learning model. The classification based on the features extracted from the histology images has shown good results in previous studies [12][13]. Hence, the detection of nuclei is crucial for correct results. Detection accuracy can be limited by variations in tissue and nuclei staining, image contrast, noisy stain blobs, overlapping nuclei, and variation in nuclei size and shape, with the latter more prominent with higher CIN grades.

In recent years, various algorithms have been proposed to segment nuclei and to extract the nuclei features from digitized medical images. The accuracy of algorithms to identify nuclei may be measured in two ways. The first measure is called nucleus detection or object-based detection. This nucleus-based scoring counts whether a ground-truth nucleus is detected or not. The second method is called nucleus segmentation, this pixel-based scoring counts accuracy pixel by pixel. Recent reviews by Xing and Yang [14] and Irshad [15] summarized techniques in this fast-evolving field for both nuclei detection and segmentation. The Irshad review provides additional material on nuclear features; the Xing and Yang review includes additional recent studies; both reviews give detailed descriptions of methods and results for nuclei detection for many types of histopathology images including brain, breast, cervix, prostate, muscle, skin and leukocyte images [7][15]. In the following, we summarize selected recent methods to find nuclei in histopathology images in general, followed by specific methods to find nuclei in cervical images.

For the general domain of histopathology images, recent studies have employed conventional techniques, various deep learning (DL) techniques, and techniques

combining both methods. A graph-cut technique was followed by multiscale Laplacian-of-Gaussian (LoG) filtering, adaptive scale selection, and a second graph-cut operation [16]. Generalized LoG filters were used to detect elliptical blob centers; watershed segmentation was used to split touching nuclei [17]. The generalized LoG filter technique was modified using directional LoG filters followed by adaptive thresholding and mean-shift clustering [18]. A convolutional neural network (CNN) nuclear detection model called “deep voting” used voting based on the location of patches and weights based on confidence in the patches to produce final nuclei locations [19]. Stacked sparse autoencoder (SSAE) DL was used for nuclei detection and compared to other DL techniques using CNN variations [20]. SSAE sensitivity was similar to that obtained for the optimal CNN; specificity compared favorably to CNN [20]. Another voting approach to overcome variable nuclear staining exploited nuclear symmetry [21]. An additional voting approach used adaptive thresholding for seed finding followed by elliptical modeling and a watershed technique [22]. Canny edge detection was followed by multi-pass directional voting; results surpassed those of the SSAE. A CNN was combined with region merging and a sparse shape and local repulsive deformable model [23] with good results.

In the domain of cervical cytology and histopathology, automated localization of the cervical nuclei used the converging squares algorithm [24]. The Hough transform was implemented to detect the nuclei based on shape features [25]. Cervical cells were classified using co-occurrence matrix (GLCM) textural feature extraction and morphological transforms [26]. Analysis of cell nuclei segmentation was performed through Bayesian interpretation after segmentation by a Viterbi search-based active

contour method [27]. Segmentation was also accomplished by a region grid algorithm through contour detection around the nuclei boundary [28]. Nuclei were segmented using level-set active contour methods [29][30]. Intensity and color information was used for nuclei enhancement and segmentation [31]. A deep learning framework was used for segmentation of cytoplasm and nuclei [32]. K-means clustering was used for nuclei feature extraction followed by classification by fusion [13]. A multi-scale CNN followed by graph partitioning was used for nuclei detection in cervical cytology images [33]. Transfer learning to recognize cervical cytology nuclei using the CaffeNet architecture was trained first on ImageNet then, using the trained network, retrained on cervical slide images, containing one cell per slide [34].

Semantic pixel-wise labeling [35] for detection of nuclei is computationally expensive, since every pixel is individually labeled through a series of encoder and decoder stacks. U-Net [36] utilizes up-sampling approach with deconvolution layers with 23 convolutional layers, which makes the network use more memory and more computations. The nuclei segmentation research here employs DL to extract nuclei patches, a simple linear iterative cluster (SLIC) model and a convolutional neural network to classify the obtained superpixel data. A group of similar pixels (superpixels) are classified, requiring reduced memory compared to the pixel-wise approach, also reducing the number of parameters to be tuned. Scoring in the current study, object-based detection, is based upon whether nuclei are correctly detected or not.

The remainder of the article is organized as follows: Section 2 (Methods) presents the image preprocessing, superpixel generation and classification used in this research;

Section 3 (Results and Analysis) presents and analyzes the results obtained; Section 4 provides the study conclusions.

## 2. METHODS

Biologically inspired CNNs operate upon a digital image, convolving image arrays with the image, producing feature vectors serving as parameters to the CNN. The automatically determined feature vectors serve as weights; these are modified with each iteration as the network learns by training.

The primary goal of this paper is to segment the nuclei in the epithelium region of cervical cancer histology images by considering local features instead of features from the whole image. This local information is used to classify whether the segment contains nuclei or background. The CNNs use image vectors as inputs and learn different feature vectors, which ultimately solve the classification problem. The proposed methodology is depicted in Figure 2.

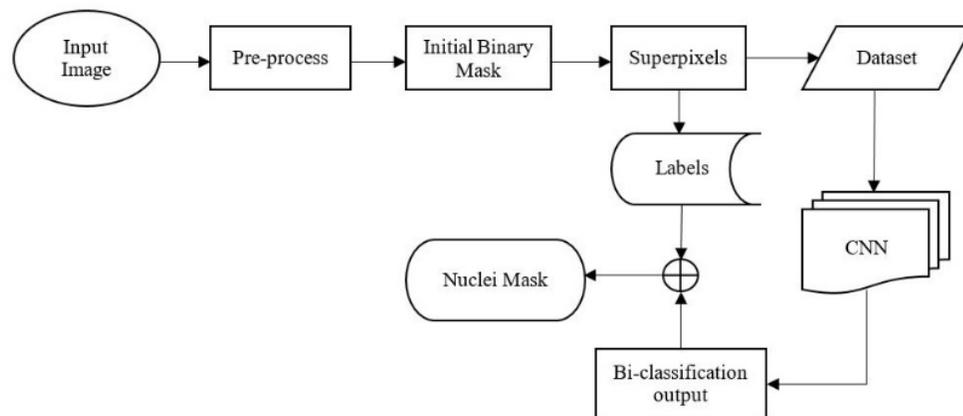


Figure 2. Proposed methodology.

In order to make use of localized information, small image patches are obtained from the original image using a superpixel extraction method. Superpixel algorithms are devised to group pixels with similar properties into regions to form clusters. Optimal superpixels avoid over segmentation without information gain, which is present at the pixel level, and under segmentation with information loss, if superpixels are too large. A simple linear iterative clustering (SLIC) algorithm is chosen as it generates superpixels based on color (intensity) and distance proximities with respect to each pixel.

## **2.1. PRE-PROCESSING**

Before extracting superpixels, the original image is preprocessed using a Gaussian smoothing filter to smooth the input image in order to reduce Gaussian and impulse valued noises, which are mainly generated during image capture from the slides and digitization process [37]. The results of over-segmented images through superpixel generation also indicate the importance of smoothing the images. The filter's impulse response is the Gaussian function, which decays rapidly, so as to select narrow windows to avoid the loss of image information. This function divides the image into its respective windows and applies the cost function. The two-dimensional Gaussian function is applied on the input image using a built-in MATLAB<sup>®</sup> function.

The standard deviation can be user-defined; here we use the default value of two. The Gaussian filter is applied instead of a trimmed mean filter because the Gaussian filter processes our images 3184.16x times faster than the trimmed mean filter. When the outputs of the algorithms were compared, the output using the Gaussian filter gave a better superpixel result compared to the output obtained using the trimmed mean filter.

The darker nuclei are in general surrounded with red stained cytoplasm inside a cell and the background region is not stained. So, the RGB color space of the image is converted to CIE LAB color space [38][33] to improve the contrast between nuclei, cytoplasm and background. The contrast is further enhanced using a linear transformation, increasing the scale of pixel intensity from  $[r_{min}, r_{max}]$  to  $[0, 255]$ . A morphological closing operation is applied on the luminance (L component) plane of the resultant CIE LAB color image to remove any small holes and to smooth boundaries. The L component represents the perceived brightness, which further increases image contrast. These operations produce the initial binary nuclei mask, to aid in extracting superpixels from the image. The generated binary mask reduces computational overload and reduces challenges due to noise and other variations in cervical histopathology images, such as variable staining present in cervical tissue, to provide a binary mask overlay to guide the next step in superpixel generation.

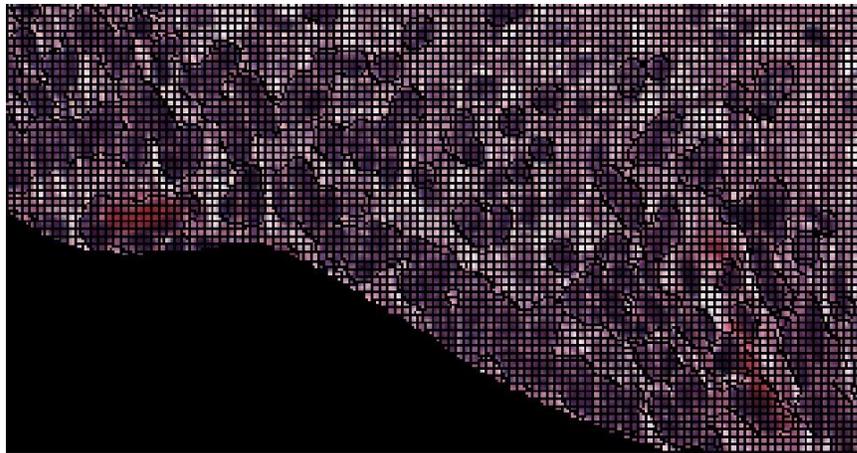


Figure 3. A portion of original image with superpixels. Nuclei do not exceed 16 pixels in height or width or 200 pixels in area.

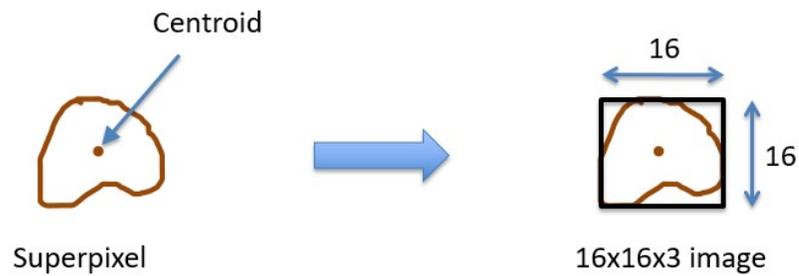


Figure 4. Generation of 16x16x3 RGB image from superpixel.

## 2.2. SUPERPIXEL EXTRACTION

Superpixels are generated automatically for the test images. A simple linear iterative clustering (SLIC) algorithm [39] is used to extract superpixels rather than other state-of-the-art methods [40][41][42][43] because it is faster, more memory efficient, has better adherence to boundaries, and improves segmentation performance. Also, it considers both color and distance properties which is appropriate with color orientation of the nuclei around a small region.

A labeled matrix, with size equal to that of the original image, is obtained as an output from the SLIC function. A manually generated epithelium mask, which is verified by an expert pathologist (RZ), is then applied on the labeled matrix to remove the unwanted region. The resultant matrix is again relabeled. The minimum size for superpixels, 200 pixels, is chosen to be larger than the largest nucleus, and smaller than the patch size (256 pixels). The patch width and height (16 pixels) are chosen to contain all superpixels and all nuclei, as shown in Figure 3, so that the whole superpixel region is covered while creating a 16x16x3 RGB patch image dataset for training the convolutional neural network.

The centroid of each superpixel is computed. With respect to that centroid, a 16x16x3 image patch is formed as shown in Figure 4. A patch is said to be a part of the nuclei region if nuclei comprise at least 10% of its area. The nuclei region is given highest priority compared to the cytoplasm and background. The problem of generating 16x16x3 patches from superpixels at the edges of the image is solved by mirroring the image.

Finally, 16x16x3 RGB input images are obtained from the superpixels of the original image. As DL benefits from more examples, data augmentation is performed by rotating the original image by 180 degrees and extracting 16x16 patches.

### **2.3. DATA GENERATION**

Data generation is done carefully to prepare both training and test image data sets. For our experiment, a total of 12 images, six images each from the 71-image dataset and 62-image dataset are used for training the network. The remaining 121 images are used in the testing phase. Thus, the training and test sets used for generating results reported in this study are disjoint. Nuclei segmentation has been investigated in previous studies using the 71-image [23] and 62-image [26] datasets, providing the benchmarks for this study. Training images are carefully chosen so that the network understands how to handle different kinds of images. Observation of images from the datasets discloses three types of images: images with light nuclei and light cytoplasm, images with darker nuclei and moderate cytoplasm, and images with darker nuclei and thicker cytoplasm as shown in Figure 5. To balance the training set for the CNN, six images from each dataset, two images for each of the three image types are included, a total of 12 images.

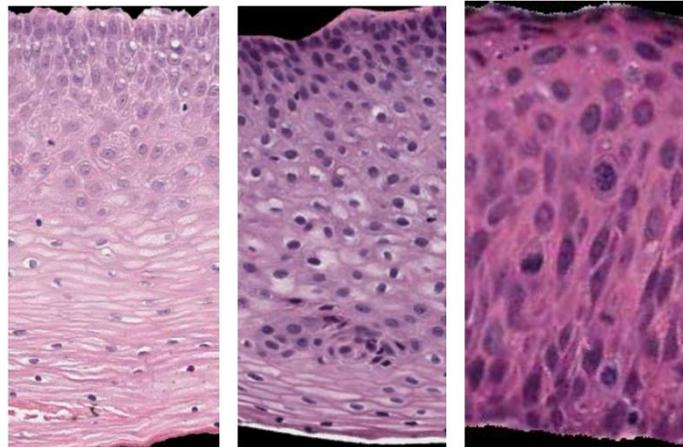


Figure 5. Images with lighter nuclei (left), darker nuclei with lighter cytoplasm (center), darker nuclei with thicker cytoplasm (right).

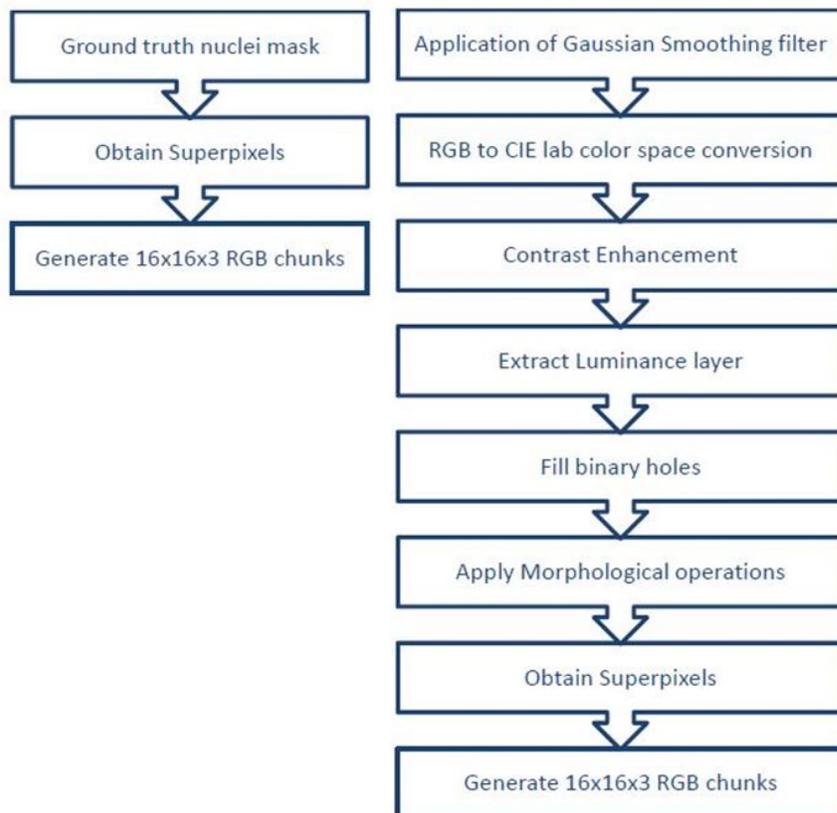


Figure 6. Generation of training dataset (left) and test dataset (right).

Classifying whether nuclei are present or not in the 16x16x3 patch is a binary classification problem. The patch target label is obtained from the binary nuclei masks that are already available in the database. Some of the portions of the nuclei masks are modified so that the target labels represent exact ground truth values. The extracted 16x16x3 patches are as shown in Figure 7. The label “0” denotes nuclei and the label “1” denotes background. A total of 377,012 patches are obtained using preprocessing steps as shown in Figure 6 (left) for 12 original images that comprise both nuclei and background.

The test data is generated by preprocessing the image (Figure 6, right). The luminance plane is used to generate superpixels, and then 16x16x3 image patches are formed for each individual original image.

## **2.4. CONVOLUTIONAL NEURAL NETWORK**

As a pre-step to train the CNN, all small image patches are converted to the HSV color plane and then the V-plane (value plane) is extracted. Before selecting the V-plane, various color planes are observed manually and are also used to train the network. The V-plane and the L-plane (luminance plane) gave promising results. The V-plane is considered for this experiment, as shown in Figure 7. The V component indicates the quantity of light reflected and is useful for extraction from the patches because the nuclei are typically blue-black and reflect only a small amount of light.

In order to classify the presence of nuclei, the CNN is trained with the features that were generated by convolutional layers using raw pixel input data. The first stage was a shallow CNN with one convolutional layer and a following max pool layer. 36,478 image patches (extracted from two images) were processed for a quick quality check. In

order to classify the presence of nuclei, the CNN is trained with the features that were generated by convolutional layers using raw pixel input data. A remarkable improvement in the validation accuracy was observed when a deep CNN network architecture (modified LeNet-5 [44] model with varied layers and hyper-parameters as shown in Figure 8) was considered with multiple convolutional, max pooling, and dropout layers at the beginning of the network and three dense neural networks (convolution and dense layers with a non-linear ReLU activation function [45]) at the end of the network. The two neurons in the output layer are activated with a SoftMax function.

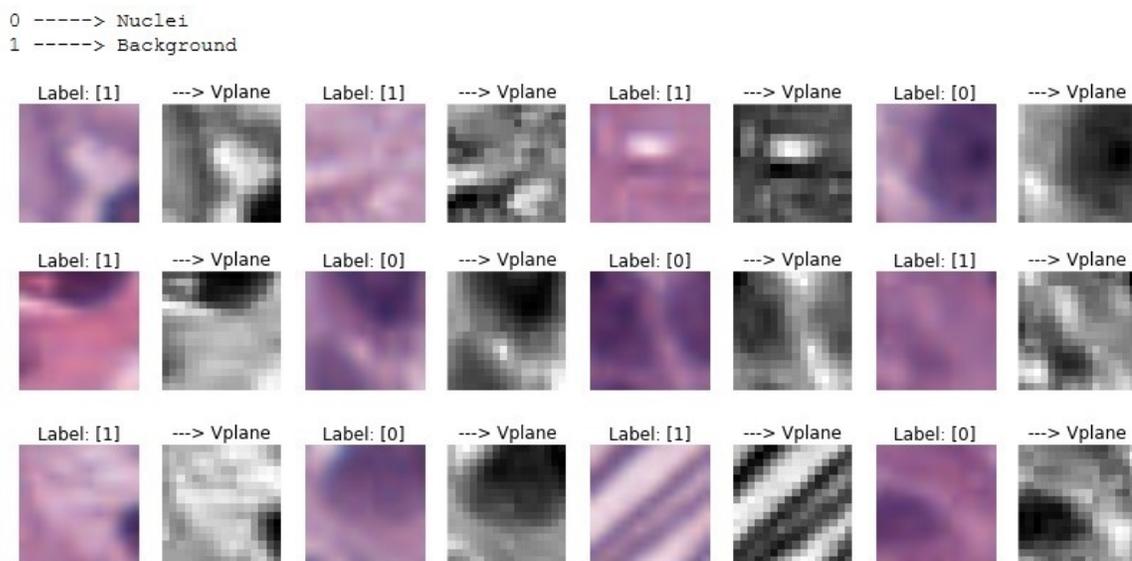


Figure 7. Sample 16 x 16 x 3 RGB images and their 16 x 16 V-plane images.

This produced 98.1% validation accuracy on two input images. Later, 10 more images were included to make the network learn to classify nuclei in different environments, as shown in Figure 5. Upon training with 377,012 patches of 16x16 size (extracted from 12 full size images), a validation accuracy of 95.70% is achieved.

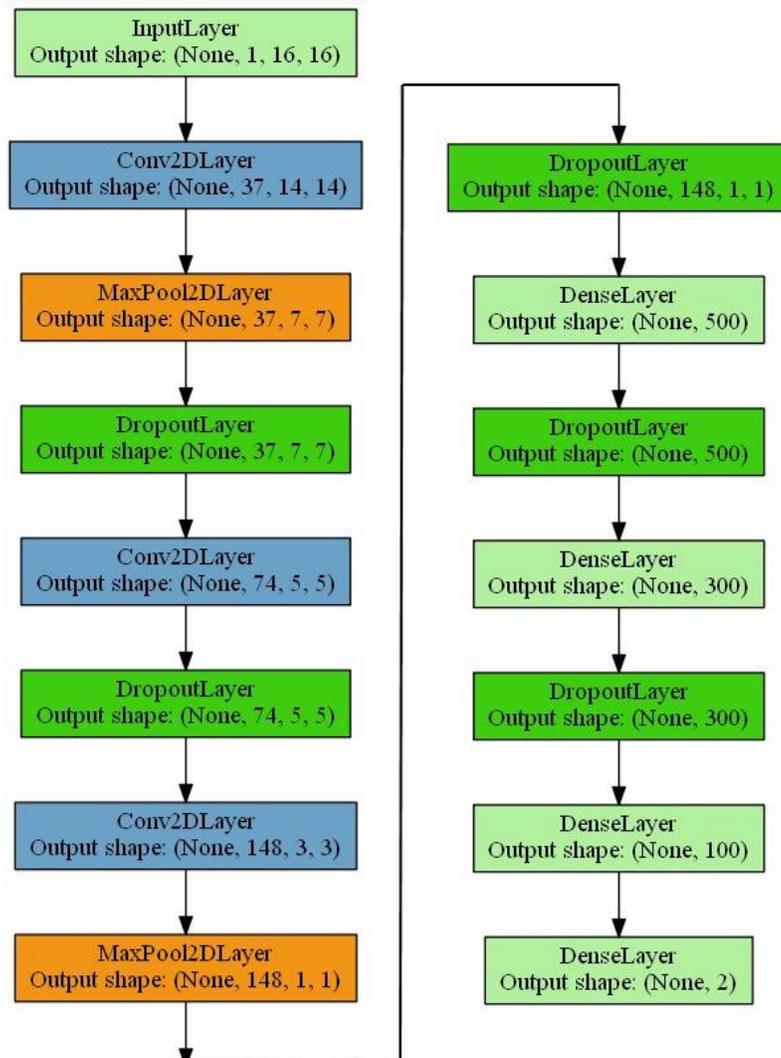


Figure 8. CNN Architecture.

The obtained data set of inputs and target labels are used to train CNNs with different architectures and the following architecture (Figure 8) gave best results with higher validation accuracy on test images that were part of the training data.

The training dataset is used to fit the CNN model. A validation dataset, consisting of 20% of the training dataset, is helpful to estimate the prediction error for best model

selection. Categorical accuracy ( $L_{\mu}$ ) is computed between targets ( $t_{i,c}$ ) and prediction ( $p_{i,c}$ ) produced from the validation dataset.

The weights are initialized randomly using Glorot weight initialization [46]. An adaptable learning rate  $\in (0.0001, 0.03)$  and momentum with range  $\in (0.9, 0.999)$  are applied to the network while training for 2000 epochs. The architecture produced a validation accuracy of 95.70% at the end of the 2000th epoch. The network is trained for 2000 epochs since further training appears not to decrease validation loss (Figure 9).

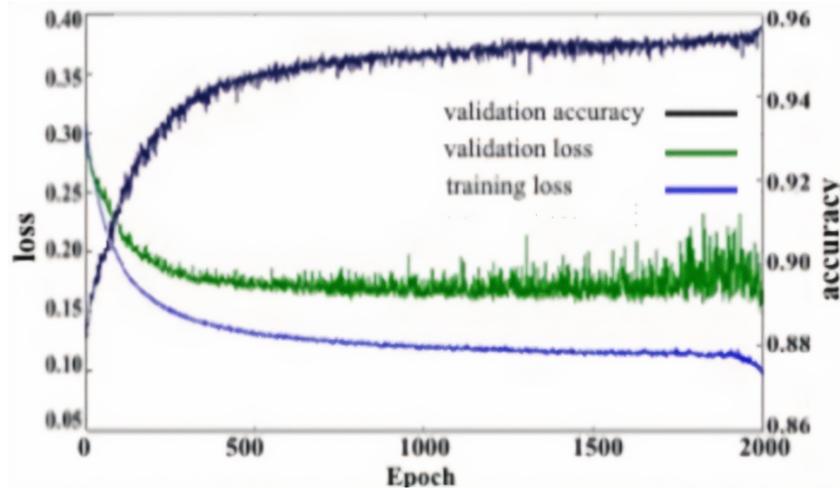


Figure 9. Training loss, validation accuracy and validation loss vs. epochs.

The error on the training set is denoted as training loss. Validation loss is the error as a result of running the validation set through the previously trained CNN. Figure 9 represents a drop-in training and validation error as the number of epochs increase. This is a clear indication that the network is learning from the data that is given as an input to the network.

Figure 10 (left) shows all 37x3x3 first-layer convolutional feature vectors obtained from the trained network. The initial layer of the convolutional network mainly focuses learning on the edge and curve features of the input image. Figure 10 (right) represents the result of the convolution of the feature vectors with the 16x16 image producing a 32x14x14 image.

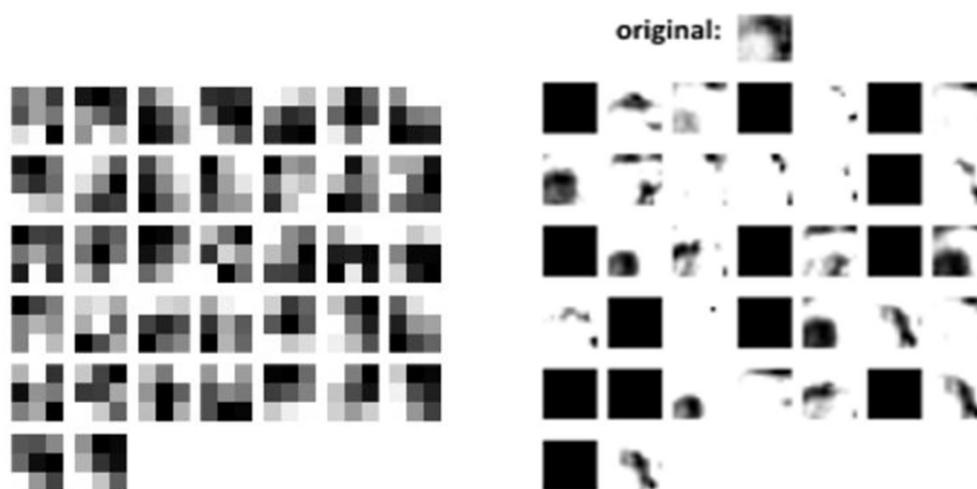


Figure 10. 32x3x3 CNN filters and 32x14x14 convolved output in first layer.

The trained network model is saved along with the weights and filter coefficients. This saved model is loaded back to test on the remaining images of the 71-image and 62-image datasets (121 images) by classifying individual patches generated from each image to assess nuclei detection accuracy. The location of every superpixel extracted from the original image is saved as a labeled image. The results of classification are mapped with the labeled image to finally obtain a binary nuclei mask from the corresponding original image. The nuclei detection rate on the test images is then calculated by manually

counting all 108,635 original ground truth nuclei truly detected and those falsely detected by the algorithm.

### 3. EXPERIMENTAL RESULTS AND ANALYSIS

#### 3.1. EXPERIMENTAL RESULTS

The proposed algorithm is applied on both 71-image, 62-image datasets., using six images from each of the datasets for training the CNN. The remaining images are used for testing the trained model. The training set and test set are disjoint. Figure 11 depicts the nuclei mask generated, with nuclei mask boundaries marked in green.

The deep learning algorithm applied to both the 71-image dataset and the 62-image dataset shows overall segmentation accuracy of 97.11% and 93.33%, respectively. Finally, the overall segmentation accuracy of the combined set is 95.97%.

The accuracy of nuclei detection is calculated on a per-nuclei basis by manually recording the True Positive (TP) (i.e., the number of nuclei successfully detected), False Negative (FN) (i.e., the number of nuclei not detected), and False Positive (FP) (i.e., number of non-nuclei objects found). Using FP and FN totals, accuracy measures are calculated [26], including precision, recall, accuracy ( $\mu$ ), dice similarity coefficient (DSC), F1 (harmonic mean of precision and recall) and Jacquard index (JAC), Equations 1-6. Table 1 shows these accuracy measures for the 62, 71 and combined datasets.

It is observed that if smaller size superpixels are considered, that is, if finer localization is done, the final nuclei masks are better. Also, a deeper CNN shows improved classification results when compared to a shallow CNN.



Figure 11. Nuclei masks (green) superimposed on the original image.

### 3.2. ANALYSIS OF RESULTS

In this section, the results from Section 3.1 are compared with results from benchmark algorithms. The following images represent the FP and FN cases. Figure 12 (left) represents a FP condition where false nuclei detection is observed. The circled portion shows the region where there is no nucleus present in the original image but detected as nucleus present with a green contour around the FP object boundary. Figure 12 (right) shows a nucleus misclassified as background. The undetected nucleus is marked in the original image, but there is no contour around the marked nucleus. Both FP and FN cases lower overall object-based detection accuracy.

Equations (4)-(9). Nuclei detection accuracy given TP and TN.

$$\text{precision} = \frac{TP}{TP + FP} \quad (1)$$

$$\text{recall} = \frac{TP}{TP + FN} \quad (2)$$

$$\mu = \frac{TP - (FP + FN)}{TP} \quad (3)$$

$$DSC = \frac{2TP}{2TP + FP + FN} \quad (4)$$

$$F1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (5)$$

$$JAC = \frac{TP}{TP + FP + FN} \quad (6)$$

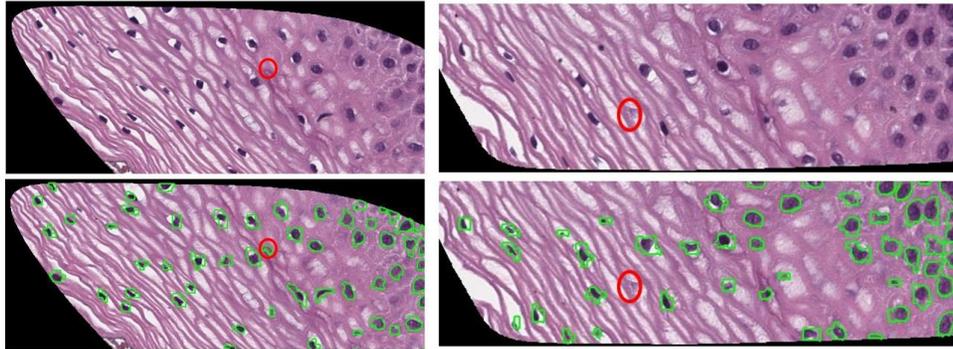


Figure 12. Examples of false positive (left) and false negative (right) results. Note variable staining.

The presence of red stains on the image samples always poses challenges in nuclei detection as the stains are falsely detected as nuclei by various algorithms; yet some nuclei may lie under red stains. The proposed algorithm has overcome this challenge by detecting the nuclei even under the red stains (Figure 13). The training process of the CNN allows learning about this feature from the ground truth images.

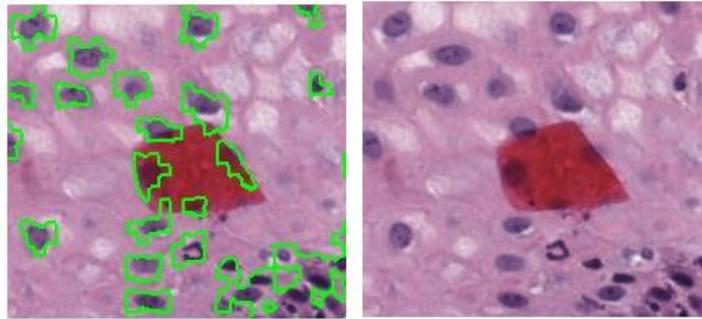


Figure 13. Nuclei detected even under red stains.

### 3.3. COMPARISON OF RESULTS

This paper presents a deep learning-based nuclei segmentation approach, using superpixel extraction followed by a CNN classifier. The algorithm has achieved an overall accuracy  $\mu$  of 96.0% on the combined set, with 97.11% accuracy achieved on the 71-image dataset (Table 1), outperforming previous cervical histopathology nuclei detection approaches. Previously, segmentation based on K-means clustering followed by mathematical morphology operations [13] produced an overall recall estimated at 89.5% on the 62-set of images. The level set method and fuzzy C-means clustering [30] approach on the 71-image dataset achieved 96.47% accuracy in comparison to the current 97.11% accuracy. Some recent results in cervical cytology nuclei detection have produced very high nuclei detection (object-based results) (Table 2) [32][47]. Nuclei detection in cervical cytology images is not comparable to nuclear detection in histopathology images. As Irshad et al. noted, nuclei segmentation “is particularly difficult on pathology images” [15]. Cervical cytology images have “well-separated nuclei and the absence of complicated tissue structures,” while most nuclei in histopathology images are “often part of structures presenting complex and irregular

visual aspects” [15]. In addition, we have found that cytology images have a greater contrast, and fewer nuclear mimics.

Table 1. Nuclei detection results using the deep learning superpixel approach.<sup>1</sup>

Data set	# Nuclei	TP	FP	TN	Precision	Recall	$\mu$	DSC	F1	JAC
71 set	75,047	74,122	925	1,218	98.76	98.38	97.11	98.57	98.56	97.19
62 set	33,588	31,928	1,660	469	95.05	98.55	93.33	96.77	96.76	93.75
Combined set	108,635	106,050	2,585	1,687	97.62	98.43	95.97	98.02	98.02	96.13

1: Accuracy measures given as percentages. TP: True positive, FP: False positive, FN: False negative, DSC: Dice similarity. coefficient, JAC: Jacquard index.

Table 2. Cervical nuclear detection vs. Current deep learning superpixel approach.<sup>1</sup>

Study	# Nuclei	TP	FP	FN	Precision	Recall	$\mu$	DSC	F1	JAC
LAGCC <sup>C</sup> [36]	420	378	67	42	85.00	90.00	71.16	87.40	87.00	77.62
MSCNN-GCC <sup>C,3</sup> [25]	33,588	-	-	-	-	-	99.00	-	-	-
CCNN <sup>C</sup> [28]	917	900	7	17	99.41	98.20	97.33	98.68	98.80	97.40
LSAC-FCM <sup>C,5,6</sup> [26]	75,107	3,791	662	1,316	97.80	98.25	95.96	98.02	98.00	96.12
<b>Current<sup>6</sup></b>	<b>75,047</b>	<b>74,122</b>	<b>25</b>	<b>1,218</b>	<b>98.76</b>	<b>98.38</b>	<b>97.11</b>	<b>98.57</b>	<b>98.56</b>	<b>97.19</b>

C: Cervical cytology study performed on pap preparations, 1: Accuracy measures given as percentages, 5: Mean of 3 test sets; TP and TN calculated from given precision and recall, 6: Current method and LSAC-FCM results for 71-set. TP: True positive, FP: False positive, FN: False negative, DSC: Dice similarity coefficient, JAC: Jacquard index.

Table 2 compares the current deep learning superpixel nuclei results with previous cervical nuclei detection studies, with results for all studies using object-based scoring. The current method outperforms the previous cervical histopathology study. Table 3 compares the current study with recent histopathology nuclei detection studies reported for various tissues, using object-based scoring. This object identification

accuracy, in comparison to pixel-based nuclear outline accuracy, may be the better of the two measures, because once a nucleus is known with high assurance, then outlines, texture and other characteristics can be scored. The current method for nuclei object detection outperforms all previous approaches.

Table 3. General nuclei detection results vs. Current deep learning superpixel approach.<sup>1</sup>

Study	# Nuclei	TP	FP	FN	Precision	Recall	$\mu$	DSC	F1	JAC
gLoG <sup>2,7</sup> [10]	13,749	11,517	1,491	2,232	88.55	83.77	67.67	86.09	86.00	75.58
Ellipse Voting <sup>2,7</sup> [15]	13,749	11,584	1,588	2,165	88.00	84.25	67.40	86.08	86.02	75.67
SSAE3 <sup>4</sup> [13]	750,000	621,375	78,051	128,625	88.84	82.85	66.74	85.74	85.74	75.04
MATDK [37]	3,381	2,979	744	402	80.02	88.11	61.54	83.87	83.87	72.22
SSDCVR- CNN <sup>5</sup> [16]	4,748	4,266	1,337	482	76.14	89.86	57.38	82.43	82.00	70.12
<b>Current</b>	<b>108,635</b>	<b>106,050</b>	<b>2,585</b>	<b>1,687</b>	<b>97.62</b>	<b>98.43</b>	<b>95.97</b>	<b>98.02</b>	<b>98.02</b>	<b>96.13</b>

1: Accuracy measures given as percentages, 2: Total number of nuclei in 21 test slides is estimated, 4: Total number of nuclei estimated from per-slide mean, 5: Mean of 3 test sets; TP and TN calculated from given precision and recall.

There has been a noticeable trend recently in the number of studies using deep learning (DL) for nuclei detection. DL is a powerful technique for nuclei detection; with sufficient numbers of nuclei, deep learning yields superior performance [14]. Yet the general enthusiasm about DL techniques should be tempered with the reality that datasets often have insufficient samples to allow learning of nuclei characteristics that vary significantly; besides nuclei size, shape, and internal features, nuclear staining varies widely [21]. Since pathologist time is a scarce resource, the number of pathologist-marked nuclei in databases remains over two orders of magnitude lower than the numbers of test nuclei in large test sets; nuclei detection results are often estimated from samples

of marked nuclei [13]. In some recent studies, detection accuracy for conventional techniques, which included incorporation of higher-level knowledge, e.g. nuclear edge symmetry, surpassed DL results (Table 3) [10][13][15][21][48].

Other studies in histopathology have surpassed deep learning results by combining conventional techniques with deep learning techniques. Zhong et al. fused information from supervised and deep learning approaches. In comparing multiple machine learning strategies, it was found that the combination of supervised cellular morphology features and predictive sparse decomposition deep learning features provided the best separation of benign and malignant histology sections [49]. Wang et al. were able to detect mitosis in breast cancer histopathology images by using the combined manually tuned cellular morphology data and convolutional neural net features [50]. Arevalo and colleagues added an interpretable layer they called “digital staining,” to improve their deep learning approach to classification of basal cell carcinoma [51]. Of interest, the handcrafted layer finds the area of interest, reproducing the high-level search strategy of the expert pathologist.

Additional higher-level knowledge has been used to separate nuclei which touch or overlap in multiple studies. However, the higher-level knowledge which pathology specialists use most extensively is the overall architecture present in the arrangement of cells and nuclei in the histopathology image. Thus, certain patterns, such as the gradient of nuclear atypia from basal layer to surface layer in carcinoma in situ, the changes as the CIN grade increases, and different patterns of a certain type of cancer, can all provide critical diagnostic information. There is an interaction between these higher-level patterns and nuclei detection; not all nuclei are of equal importance in contributing to the

diagnosis. Future studies could incorporate higher-level architectural patterns in the detection of critical cellular components such as nuclei. Thus, higher-level architectural knowledge such as nuclear distribution obtained by conventional image processing techniques fused with DL techniques will be used to advantage in automated diagnosis in the future. Since much higher-level histopathology knowledge is domain-specific, the longstanding goal of applying a single method to multiple histopathology domains remains elusive.

#### **4. CONCLUSION**

The proposed method of deep learning-based nuclei segmentation with superpixel analysis has shown improved segmentation results in comparison to state-of-the-art methods. The proposed method, over-segmenting the original image by generating superpixels, allows the CNN to learn the localized features better in the training phase. The trained model is finally applied on a larger dataset. Future work includes application of other CNN architectures as well as fusion with higher-level knowledge with the CNN classifier. Features obtained from the detected nuclei will be used in automatic CIN classification.

#### **ACKNOWLEDGEMENTS**

This research was supported [in part] by the Intramural Research Program of the National Institutes of Health (NIH), National Library of Medicine (NLM), and Lister Hill

National Center for Biomedical Communications (LHNCBC). 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 (DCEG).

## REFERENCES

- [1] Mcauliffe MJ, Lalonde FM, MCGarry D, Gandler W, Csaky K, Trus BL. Medical Image Processing, Analysis & Visualization In Clinical Research, pp. 381–386.
- [2] Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012, *Int. J. Cancer*. 2015;136(5):E359–E386.
- [3] He L, Long LR, Antani S, Thoma GR. Histology image analysis for carcinoma detection and grading. *Comput Methods Programs Biomed.*,2012;107(3):538–556.
- [4] Kumar V, Abba A, Fausto N, Aster J. The female genital tract in Robbins and Cotran Pathologic Basis of Disease, 9th ed., 2014; chapter. 22.
- [5] He L, Long LR, Antani S, Thoma GR, Computer assisted diagnosis in histopathology in *Sequence and Genome Analysis: Methods and Applications*, iConcept Press. Hong Kong, 2011;15:271–287.
- [6] Wang Y, Crookes D, Eldin OS, Wang S, Hamilton P, Diamond J. Assisted diagnosis of cervical intraepithelial neoplasia (CIN). *IEEE J Sel Top Signal Process*. 2009;3(1):112–121.
- [7] Egner JR. AJCC Cancer Staging Manual, *JAMA*. 2010;304(15):1726.
- [8] He L, Long LR, Antani S, Thoma GR. Computer assisted diagnosis in histopathology. *Seq Genome Anal Methods Appl*. 2010;3:271–287.
- [9] Wang Y, Crookes D, Eldin OS, Wang S, Hamilton P, Diamond J, Assisted diagnosis of cervical intraepithelial neoplasia (CIN), *IEEE J Sel Top Signal Process*. 2009;3(1):112–121.

- [10] McCluggage WG, Walsh MY, Thornton CM, Hamilton PW, Date A, Caughley LM, Bharucha H. Inter- and intra-observer variation in the histopathological reporting of cervical squamous intraepithelial lesions using a modified Bethesda grading system. *Br J Obstet Gynaecol.* 1998;105(2):206–10.
- [11] Ismail SM, Colclough AB, Dinnen JS, Eakins D, Evans DM, Gradwell E, O’Sullivan JP, Summerell JM, Newcombe R. Reporting cervical intra-epithelial neoplasia (CIN): intra- and interpathologist variation and factors associated with disagreement. *Histopathology,* 1990;16(4):371–376.
- [12] De S, Stanley RJ, Lu C, Long R, Antani S, Thoma G, Zuna, R. A fusion-based approach for uterine cervical cancer histology image classification. *Comput Med Imaging Graph.* 2013;37(7-8):475–87.
- [13] Guo P, Banerjee K, Stanley RJ, Long R, Antani S, Thoma G, Zuna R, Frazier SR, Moss RH, Stoecker WV. Nuclei-based features for uterine cervical cancer histology image analysis with fusion-based classification. *IEEE J Biomed Health Inform.* 2016;20(6):1595-1607.
- [14] Xing F, Yang L. Robust nucleus/cell detection and segmentation in digital pathology and microscopy images:A comprehensive review. *IEEE Rev Biomed Eng.* 2016;9:234-263.
- [15] Irshad H, Veillard A, Roux L, Racoceanu D. Methods for nuclei detection, segmentation and classification in digital histopathology: A review—Current status and future potential. *IEEE Rev Biomed Eng.* 2014;7:97–114.
- [16] Al-Kofahi Y, Lassoued W, Lee W, Roysam B. Improved automatic detection and segmentation of cell nuclei in histopathology images. *IEEE Trans Biomed Eng.* 2010;57(4):841–852.
- [17] Kong H, Akakin HC, Sarma SE. A generalized Laplacian of Gaussian filter for blob detection and its applications. *IEEE Trans Cybern.* 2013;43(6):1719–1733.
- [18] Xu H, Lu C, Berendt R, Jha N, Mandal M. Automatic nuclei detection based on generalized Laplacian of Gaussian filters. *IEEE J Biomed Health Inform.* 2017;21(3):826-837.
- [19] Xie Y, Kong X, Xing F, Liu F, Su H, Yang L. Deep voting: A robust approach toward nucleus localization in microscopy images, in *Proceedings of 18th International Conference on Medical Image Computing and Computer-Assisted Intervention -- MICCAI; 2015 Oct 5-9; Munich, Germany. Proceedings, Part III.* Navab N, Hornegger J, Wells WM, Frangi AF, eds. Berlin, Springer: 374–382.
- [20] Xu J, Xiang L, Liu Q, Gilmore H, Wu J, Tang J, Madabhushi A. Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images. *IEEE Trans Med Imaging.* 2016;35(1):119–130.

- [21] Lu C, Xu H, Xu J, Gilmore H, Mandal M, Madabhushi A. Multi-pass adaptive voting for nuclei detection in histopathological images. *Sci Reports*. 2016; 62016;6:33985.
- [22] Lu C, Mandal M. An efficient technique for nuclei segmentation in histopathological images based on morphological reconstructions and region adaptive threshold. *Pattern Recognit Lett*. 2014;18(5):1729–1741.
- [23] Xing F, Xie Y, Yang L. An automatic learning-based framework for robust nucleus segmentation. *IEEE Trans Med Imaging*. 2016;35(2):550–566.
- [24] Ricketts IW, Banda-Gamboa H, Cairns AY, Hussein K, Hipkiss W, McKenna S, Parianos E. Towards the automated prescreening of cervical smears, in *IEE Colloquium on Applications of Image Processing in Mass Health Screening*; 1992 Oct 11; London, UK. p. 7/1-7/4.
- [25] Thomas ADH, Davies T, Luxmoore AR. The Hough transform for locating cell nuclei, *Anal Quant Cytol Histol*. 1992;14(4):347–353.
- [26] Walker RF, Jackway P, Lovell B, Longstaff ID. Classification of cervical cell nuclei using morphological segmentation and textural feature extraction, *Proceedings Australia New Zealand Intell. Inf. Syst. Conf. ANZIS*; 1994; Brisbane, Australia, pp. 297-301.
- [27] Bamford P, Lovell B. Bayesian analysis of cell nucleus segmentation by a Viterbi search based active contour. *Proceedings 14th Int. Conf. Pattern Recognit*; 1998 Aug 11-20; Brisbane, Australia. vol. 1, pp: 1–3.
- [28] Krishnan N, Nandini Sujatha SN. A fast geometric active contour model with automatic region grid, 2010 *IEEE Int. Conf. Comput. Intell. Comput. Res. ICCIC*; 2010 28-29 Dec; Coimbatore, India. *IEEE 2010*, pp. 361–368.
- [29] Fan J, Wang R, Li S, Zhang C. Automated cervical cell image segmentation using level set based active contour model. *12th Intn. Conf. Control Autom. Robot. Vis. (ICARCV)*; 2012 5-7 Dec; Guangzhou, China. vol. 2012, pp. 877–882.
- [30] Edulapuram R, Stanley RJ, Long R, Antani S, Thoma G, Zuna R, Stoecker WV, Hagerty J. Nuclei segmentation using a level set active contour method and spatial fuzzy C-means clustering. *Proceedings 12th International Conference on Computer Vision Theory and Applications VISIGRAPP 2017*; Feb 27- Mar 1, Porto, Portugal. pp. 195-202.
- [31] Guan T, Zhou D, Fan W, Peng K, Xu C, Cai X. Nuclei enhancement and segmentation in color cervical smear images, 2014 *IEEE Int. Conf. Robot. Biomimetics, IEEE ROBIO 2014*; Dec 5-10, Bali, Indonesia. no. 61375032, pp. 107–112.

- [32] Song Y, Zhang L, Chen S, Ni D, Li B, Zhou Y, Lei B, Wang T. A deep learning based framework for accurate segmentation of cervical cytoplasm and nuclei Conference Proceedings 35th Annual International Conference IEEE Engineering in Medicine and Biology Society; 2014 Aug 26-30; Chicago IL. 2014:2903–2906.
- [33] Song Y, Zhang L, Chen S, Ni D, Lei B, Wang T. Accurate segmentation of cervical cytoplasm and nuclei based on multiscale convolutional network and graph partitioning, *IEEE Trans Biomed Eng.* 2015;62(10):2421–2433.
- [34] Zhang L, Lu L, Nogues I, Summers R, Liu S, Yao J. DeepPap: Deep convolutional networks for cervical cell classification. *IEEE J Biomed Health Inform.* 2017;21(6):1633-1643.
- [35] Badrinarayanan V, Handa A, Cipolla R. SegNet: A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling. arXiv:1505.07293v1 2015. accessed at <https://arxiv.org/pdf/1505.07293.pdf> accessed January 15, 2018.
- [36] Ronneberger O, Fischer P, and T. Brox T. U-Net: Convolutional networks for biomedical image segmentation. Proceedings Medical Image Computing and Computer-Assisted Intervention -- MICCAI 18th International Conference; 2015 Oct 5-9; Munich, Germany. Proceedings, Part III, Navab N, Hornegger J, Wells WM, Frangi AF. Eds. Cham: Springer International Publishing, 2015, pp. 234–241.
- [37] Kumar Boyat A, Joshi BK, A review paper: Noise models in digital image processing. *Signals Image Process: An Int Journal (SPIPJ)* 2015;6(2):63-75. Available at <https://arxiv.org/ftp/arxiv/papers/1505/1505.03489.pdf> accessed Jan 16, 2018.
- [38] Hill B, Roger T, Vorhagen FW. Comparative analysis of the quantization of color spaces on the basis of the CIELAB color-difference formula. *ACM Trans Graph.* 1997;16(2):109-154.
- [39] Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Süsstrunk S. SLIC superpixels compared to state-of-the-art superpixel methods, *IEEE Trans Pattern Anal Mach Intell.* 2012;34(11): 2274–2281.
- [40] Boykov YY, Jolly MP. Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images. *Computer Vision, 2001. ICCV 2001. Proceedings of the 8<sup>th</sup> IEEE Int Conference on Computer Vision; 2001 Jul; Vancouver, Canada.* pp:105–112.
- [41] Shi J, Malik J. Normalized cuts and image segmentation, *IEEE Trans Pattern Anal Mach Intell.* 2000;22(8):888–905.
- [42] Comaniciu D, Meer P. Mean shift: A robust approach toward feature space analysis, *IEEE Trans. Pattern Anal Mach Intell.* 2002;24(5):603–619.

- [43] P. F. Felzenszwalb PF, and D. P. Huttenlocher DP. Efficient graph-based image segmentation. *Int J Comput Vis.* 2004;59(2):167–181.
- [44] LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998; 86(11):2278-2324.
- [45] Nair V, Hinton GE. Rectified linear units improve restricted Boltzmann machines. *Proceedings 27th Int Conf Mach Learn.* 2010 Jun 21-24; Haifa, Israel. Omnipress; 2010.
- [46] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks, in *Proceedings of the 13<sup>th</sup> International Conference on Artificial Intelligence and Statistics*; 2010 May 13-15; Sardinia, Italy. vol. 9, pp. 249–256.
- [47] Zhang L, Kong H, Chin CT, Liu S, Chen Z, Wang T, Chen S. Segmentation of cytoplasm and nuclei of abnormal cells in cervical cytology using global and local graph cuts. *Comput Med Imaging Graph.* 2014;38(5):369–380.
- [48] Lu C, Mahmood M, Jha N, Mandal M. A robust automatic nuclei segmentation technique for quantitative histopathological image analysis. *Anal Quant Cytol Histol.* 2012;34,296–308.
- [49] Zhong C, Han J, Borowsky A, Parvin B, Wang Y, Chang H. When machine vision meets histology: A comparative evaluation of model architecture for classification of histology sections. *Med Image Anal.* 2017;35:530-543.
- [50] Wang H, Cruz-Roa A, Basavanhally A, Gilmore H, Shih N, Feldman M, et al. Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features. *J Med Imaging*, 2014;3:034003.
- [51] Arevalo J, Cruz-Roa A, Arias V, Romero E, González FA. An unsupervised feature learning framework for basal cell carcinoma image analysis. *Artif Intell Med.* 2015;64:131-45.

## II. EPITHNET: DEEP REGRESSION FOR EPITHELIUM SEGMENTATION IN CERVICAL HISTOLOGY IMAGES

Sudhir Sornapudi<sup>1</sup>, Jason Hagerty<sup>1,5</sup>, R. Joe Stanley<sup>1</sup>, William V. Stoecker<sup>5</sup>, Rodney Long<sup>2</sup>, Sameer Antani<sup>2</sup>, George Thoma<sup>2</sup>, Rosemary Zuna<sup>3</sup>, Shellaine R. Frazier<sup>4</sup>

<sup>1</sup>Department of Electrical and Computer Engineering, Missouri University of Science and Technology, Rolla, MO

<sup>2</sup>Lister Hill National Center for Biomedical Communications for National Library of Medicine, National Institutes of Health, Bethesda, MD

<sup>3</sup>Department of Pathology, University of Oklahoma Health Sciences Center, Oklahoma City, OK

<sup>4</sup>Surgical Pathology Department, University of Missouri Hospitals and Clinics, Columbia, MO

<sup>5</sup>Stoecker & Associates, Rolla MO

### ABSTRACT

Automated pathology techniques for detecting cervical cancer at the premalignant stage have advantages for women in areas with limited medical resources. This article presents EpithNet, a deep learning approach for the critical step of automated epithelium segmentation in digitized cervical histology images. EpithNet employs three regression networks of varying dimensions of image input blocks (patches) surrounding a given pixel, with all blocks at a fixed resolution, using varying network depth. The proposed model was evaluated on 311 digitized histology epithelial images and the results indicate that the technique maximizes region-based information to improve pixel-wise probability estimates. EpithNet-mc model, formed by intermediate concatenation of the convolutional layers of the three models, was observed to achieve 94% Jaccard index

(intersection over union) which is 26.4% higher than the benchmark model. EpithNet yields better epithelial segmentation results than state-of-the-art benchmark methods.

## 1. INTRODUCTION

In recent years, the number of cervical cancer cases worldwide has increased, making it the fourth most frequent cancer in women. It is estimated that a total of 570,000 new cases were reported in 2018, 6.6% of all female cancers. Low- and middle-income countries account for 90% of deaths from cervical cancer [1]. Prevention of cervical cancer mortality is possible with earlier treatment through screening and earlier diagnosis at the pre-cancer stage. The standard diagnostic process is the microscopic evaluation of histology images by a qualified pathologist [2][3]. The severity of cervical pre-cancer typically increases as the immature atypical cells increase across the epithelium region. Based on this observation, the pre-cancer condition affecting squamous epithelium is classified as normal or three grades of cervical intraepithelial neoplasia (CIN): CIN1, CIN2, and CIN3 [4][5][6]. Normal means there is no dysplasia and CIN1, CIN2, CIN3 correspond to mild, moderate and severe dysplasia, respectively. As the severity of the dysplasia increases, an increase in the density of immature atypical cells can be observed from lamina propria (region below epithelium) to the outer layer of epithelium. Figure 1 shows the cervical histology digital microscopy (DM) image at 10X magnification containing background, stratified squamous epithelium and lamina propria; with the epithelial binary mask (right) determined manually by a pathologist.

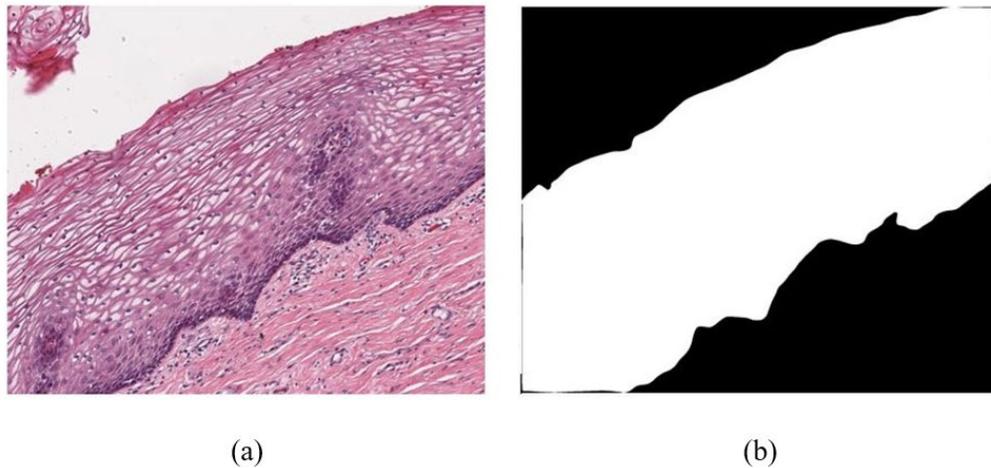


Figure 1. (a) DM image at 10X magnification with corresponding (b) manually generated mask.

Pathologists examine the epithelial regions of the cervical histology slides under a light microscope after a biopsy. The regions of lamina propria and background, which occupy the majority of the image area, are not the regions of interest during the analysis. The whole slide digital microscopy (DM) image is usually an ultra-large image, up to 40K x 80K pixels. This makes the manual examination of the DM image and segmentation of the epithelium region a tedious job. As a future step toward segmenting the epithelium in the whole slide image, we work with the higher resolution sub-images containing epithelium that share the borders with lamina propria and background as depicted in Figure 1(a). These sub-images have been cropped from the whole slide image by the pathologist. Figure 2 shows the epithelium analysis process that has been explored in previous research [7][8] using manual epithelium region segmentation. The segmented epithelium regions were split into multiple vertical segments with reference to the detected medial axis. Each vertical segment is processed to extract a set of 27 features which are later categorized into a CIN grade by applying traditional machine learning

algorithms. All the predicted CIN grades were fused through a voting scheme to generate a single CIN grade representing the entire image. The fusion based CIN grades were evaluated against the labels provided by the expert pathologist.

The goal of this research is to automate the epithelium analysis process. The primary step that needs automation is segmentation of epithelium regions to facilitate computer-assisted feature and CIN classification to assist the pathologist in the diagnostic process.

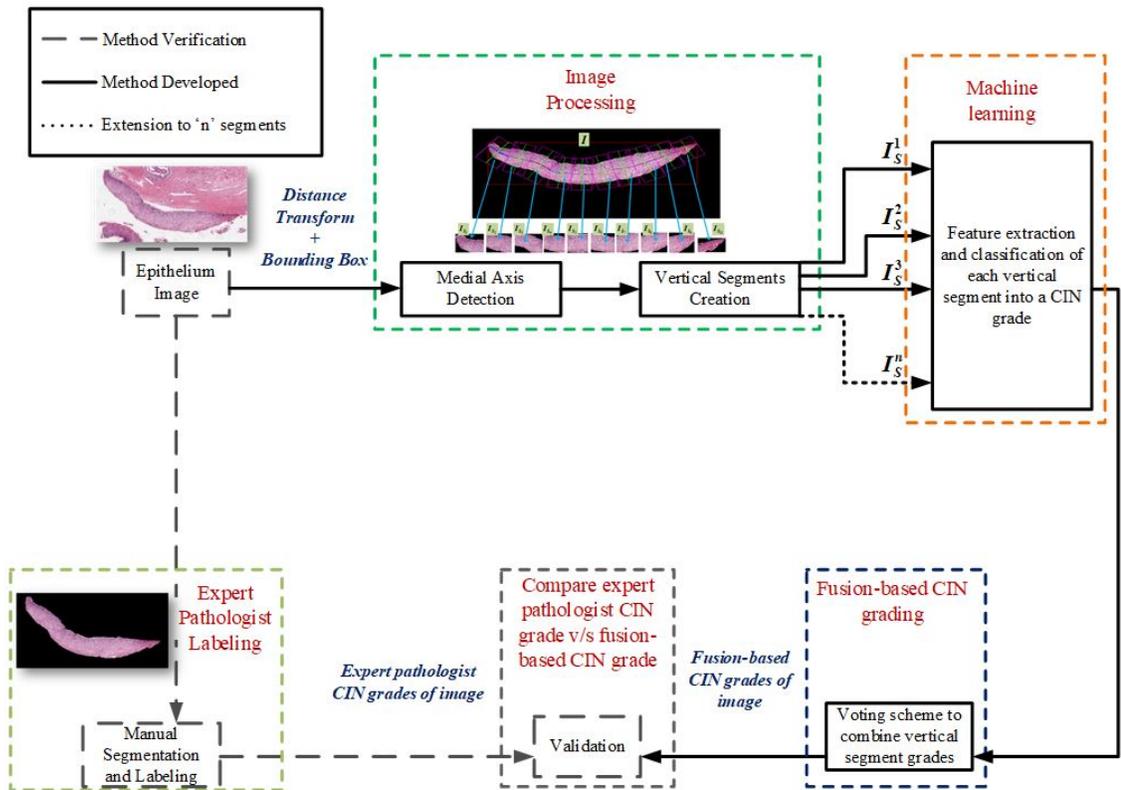


Figure 2. Epithelium analysis process used in previous research based on a manually segmented epithelium.

In this paper, we propose an automated segmentation of epithelium regions at high resolution of 10X magnification histology images, which can be applied to accurate segmentation of epithelium regions in both high-resolution and low-resolution images.

We explore the possibility of constructing small-scale but efficient convolutional neural networks (CNNs) to solve the difficult automated segmentation task. The task is challenging due to varying levels of Hematoxylin and Eosin (H&E) staining, the varying shapes of epithelial regions, the varying density and shape of cells in these regions, the presence of blood in the tissue sample and the presence of columnar cellular regions. CNNs extract hierarchal features, which contain information about patterns, colors, textures, etc. These features help the model to better predict a pixel-wise probability of a pixel belonging to the epithelium region. We design a CNN regression model that can analyze the spatial information around a pixel in the form of input image data and learn the features to assign a probability value of being epithelial pixel.

In the last decade various papers have been published on the epithelium segmentation topic with the help of conventional image processing techniques. A multi-resolution segmentation strategy [9] was developed to segment squamous epithelial layer in virtual slides. The segmentation was initially performed on a low-resolution image and later tuned at higher resolution of 40X magnification by utilizing an iterative boundary expanding-shrinking method. This is a block segmentation approach implemented with a Support Vector Machine (SVM) classifier using textural features of the image. This work was further extended [5] to diagnose CIN from the changes of density of nuclei along the perpendicular line feature. Feature-based automated segmentation was proposed [10] to segment pan-cytokeratin stained histology images of lung carcinoma by extracting

superpixels. The results were analyzed using leave-one-out methodology and achieved a Dice coefficient score of 91% for vital tumor and 69% for necrosis. Local binary patterns [11] were analyzed for precise and better segmentation of image samples from video content of respiratory epithelium. U-net [12] is a popular deep learning approach for biomedical image segmentation that is successful in segmenting various biomedical images, which we use for benchmark performance comparison in this study, where we investigate creation of an epithelium probability mask through regression analysis using a deep learning (DL) framework.

## 2. METHODS

The proposed method of epithelium segmentation is based on the idea of estimating the probability that a given pixel represents epithelium. The rationale for the probabilistic model is that, unlike segmentation of more defined biological samples such as the heart or liver, which have a discrete boundary, microscopic tissue segmentation boundaries marked by pathologists can vary significantly. A neighborhood of  $n \times m$  pixels centered on the pixel of interest to be passed to a CNN model is shown in Figure 3. The resultant scalar represents the probability that the pixel at the center of the given neighborhood belongs to epithelium. This continues in sliding-window fashion until each pixel in the digital epithelium image is processed. The final output is a probability map.

## **2.1. DATA**

The dataset for this research consisted of 351 high resolution DM histology color images and corresponding manually segmented epithelial layer masks as shown in Figure 1. The manually generated masks were verified and approved by expert pathologists, including 40 histology images representing the diversity of the data (10 images from each CIN class). The cervical histology images have varying density of nuclei and cytoplasm. There are images with dark and larger nuclei with thick cytoplasm, images with dark and smaller nuclei with moderate cytoplasm, and images with light and relatively moderate size nuclei with light cytoplasm. Also varying CIN grades show varying nuclei densities in the epithelium regions. These 40 images were employed for training the model and the remaining 311 images were used for testing and evaluating the performance of the model.

## **2.2. OVERVIEW OF PROPOSED SEGMENTATION METHOD**

The proposed epithelial segmentation task is split into four parts: 1) Data preprocessing, 2) Training, 3) Testing, and 4) Post-processing. The data preprocessing is the first step that deals with generation of smaller patch image data and normalizing the data. Section 2.3 highlights the details. Training and Testing include creating a regression CNN model and usage of memory optimized workflow in the testing phase. Section 2.4 provides more details about the architecture and workflow. Post-processing includes thresholding and generating a binary mask which is further cleaned and smoothed over the edges. Section 2.5 provides insights about the post-processing steps.

### 2.3. INPUT IMAGE DATA

CNNs needs voluminous image data of a standard shape as input. The limited availability of annotated data in this domain, consisting of 351 images of varying sizes here, is a major challenge. The dissimilar size issue could be rectified by resizing images to a standardized size, but this may introduce problems relating to cropping, aspect ratio and padding. Even if the dissimilar image sizes were not an issue, the small number of image samples is. The solution to the small dataset and dissimilar image sizes was to decompose each image into a set of overlapping patches with a patch stride  $s$ . An epithelial image of size  $(N, M)$ , would generate  $P$  patches (Equation 1).

$$P = \left\lfloor \frac{N-n+s}{s} \right\rfloor \left\lfloor \frac{M-m+s}{s} \right\rfloor \quad (1)$$

The image data is decomposed using Equation 1, with patch size  $(n, m) =$

$\{(16,16), (32,32), (64,64)\}$  and stride  $s = 16$  for training data and  $s = 4$  for test data.

This way the original RGB image and the binary ground truth masks are decomposed into a set of  $n \times m \times 3$  and  $n \times m \times 1$ , respectively over-lapping patches.

A training image dataset was created by considering 40 images representative of the four CIN grades. These images were chosen such that the network could learn various characteristic features of the histology images with different shape, color and density of nuclei and cytoplasm in both epithelial and lamina propria regions. A total of 254,514 image patches of size  $n \times m \times 3$  were generated, with 85% of the data used for the training dataset and the remaining data used to validate the trained model.

A CNN is used to solve the regression problem by predicting the probability of each pixel of the image belonging to the epithelium region. The ground truth patches were further reduced to a numerical representation of the percentage of non-zero pixels

within a given mask patch as shown in Figure 3. If the non-zero pixels in a mask patch,  $p_{mask}$ , were assigned a value of 1, the average epithelium density,  $\mu$ , of each patch is given by

$$\mu = \frac{1}{mn} \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} p_{mask}(x, y) \quad (2)$$

The ground-truth probability value for each patch is defined by the  $\mu$  value.

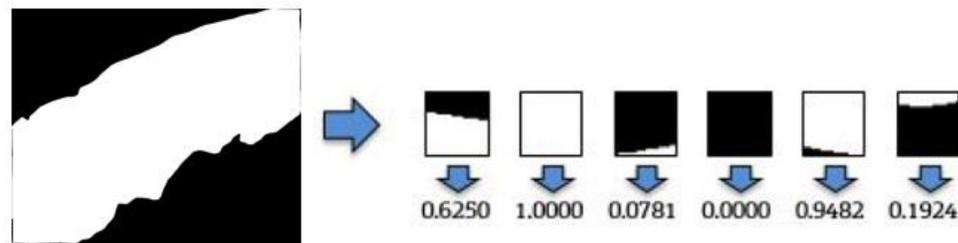


Figure 3. Generation of labels.

## 2.4. REGRESSION MODEL

Next, a regression model is determined using a CNN to predict the average epithelium density (used interchangeably with probability) in each patch image. CNNs are mostly used to classify images. The classification task has a discrete output. However, to predict probability on neighboring blocks of images, we consider the neighboring blocks as continuous data that can be handled better through a regression model and hence, we include a single neural node at the end of the network.

We design CNN models that are variants of a VGG network [13] in terms of filter's receptive field and depth. These models are fed with block images (RGB) of sizes  $16 \times 16$ ,  $32 \times 32$  or  $64 \times 64$  as shown in Figure 4. The models are named as EpithNet-16, EpithNet-32 and EpithNet-64, where the postscript represents the size of the input

image that the model can read. Each image is subjected to a stack of convolutional layers (Conv), where the first layer filter has a receptive field of  $5 \times 5$  and the following layers are designed to have  $3 \times 3$  filters. Spatial padding is applied such that the output layer has size same as the input layer after the convolution operation. The stride is fixed to 1. Each convolutional layer is followed by a  $2 \times 2$  max-pooling layer. A series of convolutional and max-pooling layers are stacked with increasing feature depth until a layer of size  $4 \times 4$  is obtained. These are then followed by four fully connected (FC) layers. The first layer has 4096 channels, followed by two FC layers containing 512 channels, and finally by a regression layer implemented as a single node FC layer. For activation functions, the Conv layers are implemented with ReLU [14], FC layers with leaky ReLU [15] and the output layer with tanh. Dropout layers were included to regularize the model to avoid overfitting.

The model is trained with augmented data. The input data is randomly augmented with shear range varying from 0 to 10 and random rotation of images between 0 to 90 degrees. The model is compiled with the Adadelata optimizer [16], which adapts the learning rate based on gradient updates. The learning rate is set to 1.0 and the gradient decay factor at each time step is set to 0.95. The loss functions investigated include L1 loss, L2 loss, log-cosh loss, normalized exponential loss, weighted Gaussian loss, and mean weighted Gaussian loss. The model is observed to perform better with L1 loss (mean absolute error) as cost function. Validation data is used to auto-tune the hyper-parameters in the network. The network is allowed to train for 300 epochs with early stopping.

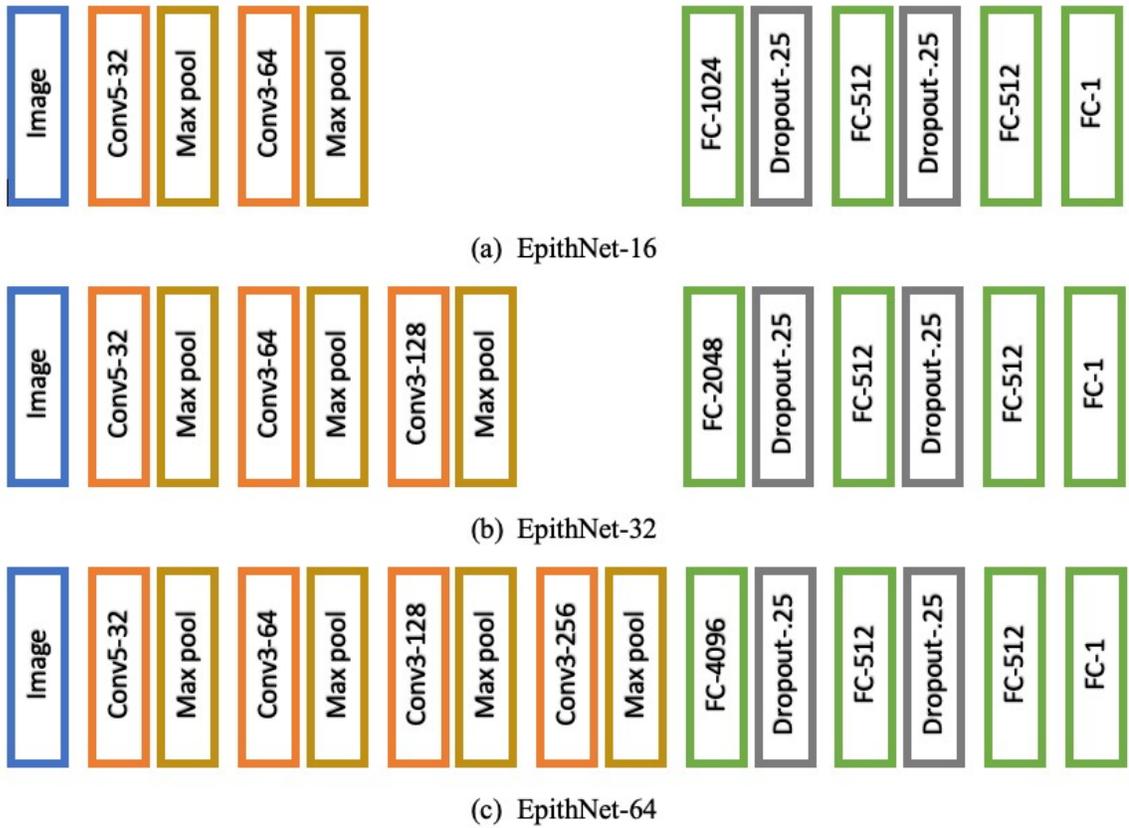


Figure 4. EpithNet architecture.

During prediction on test data, the generation of patches with respect to each pixel as centroid from the entire image poses a major challenge for memory requirements. To address this problem, we sliced the original image into smaller tiles such that each tile has approximate size  $p \times q$ . The number of tiles,  $n_t$  from  $M \times N$  image is calculated from

$$n_t = \left\lceil \frac{MN}{pq} \right\rceil. \text{ The image is split into } n_r = \left\lceil \sqrt{n_t} \right\rceil \text{ rows and } n_c = \left\lceil \frac{n_t}{n_r} \right\rceil \text{ columns of tiles.}$$

Typically, we choose  $(p, q) = (400, 400)$  for our experiments to handle memory problems. Before dividing the image into  $n_r \times n_c$  tiles, we padded the edges of the image to mirror the pixel values for a uniform split of the image into tiles. The amount of padding is calculated from  $pad_c = n_c s - rem(N, n_c s)$  across the width of the image and

$pad_r = n_r s - rem(M, n_r s)$  across the height of the image. Half of each padding rows and columns are distributed on either side of the image. The resultant image is split into  $n_t$  tiles as shown in Figure 5. Each of these smaller tile images are considered one by one to generate patch images with stride  $s = 4$  (chosen empirically without performance degradation) and the individual patches are tested through the regression model. The generated confidence value of each pixel centroid is reshaped to obtain the mask of the corresponding section of the image. This process is repeated by clearing the local memory of the patches once the mask is generated. The output mask is as shown in Figure 6.

The generated mask tiles are later stitched using a reference label image generated during the splitting process. The resultant output mask is resized by a factor of 4 (equivalent to stride  $s = 4$ ) to match the size of original input image. The output fuzzy mask is processed further to obtain a clean binary epithelial segmentation mask.

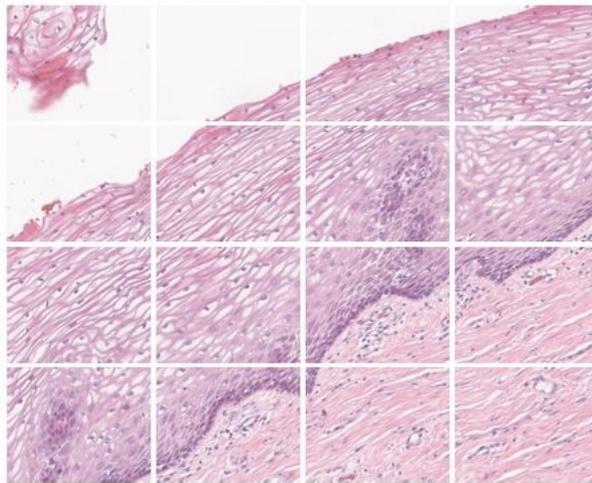


Figure 5. Original image split into tiles.

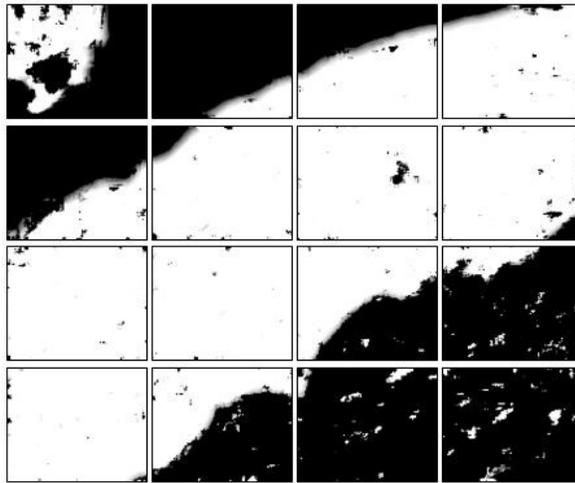


Figure 6. Predicted mask from each tile of original image.

## 2.5. POST-PROCESSING

Post-processing includes the removal of unwanted noise in the mask and smoothing the edges of the segmentation. In the epithelial mask generation, there is always a problem of drawing an exact boundary even by an expert pathologist. Considering this situation, we created a model that can generate a gradient mask; this gives us ability to choose an appropriate threshold that can satisfy the pathologist conditions. We choose values between 0.35 and 0.5 as an optimal threshold range, obtained empirically. By default, we consider 0.5 as our threshold. Since the epithelial region covers most of the image area, thresholding with this value is applied to retain the object with maximum area in the image; the remaining image area is masked as background.

The edges in the mask appear to be abruptly changing as shown in Figure 7(a) and smoothing of the edge contour is accomplished by approximating a Bezier curve. This is a parametric curve controlled by Bezier control points. The Bernstein polynomial forms

the basis of the curve. We converted the contour of the segmentation mask, which is a continuous curve, into point data. The end of the point data is appended with the first two data points which helps in closing the curve smoothly. The mid-points for every set of adjacent points which helps in closing the curve smoothly. The mid-points for every set of adjacent points were calculated and included in the data points. The updated point data is of length  $2(r + 1) + 1$ . A quadratic Bezier curve [17] is approximated by plotting a piece-wise continuous curve using three sets of control points iteratively using Equation 3.

$$B(t) = P_{i+1} + (1 - t)^2(P_i - P_{i+1}) + t^2P_{i+2} \quad (3)$$

where  $t \in [0,1]$  and  $i \in [0,2(r + 1) - 1]$ . The final curve is converted into a binary mask as shown in Figure 7(b). The resultant output mask covers the entire epithelium region and segments the region with high accuracy. This can be clearly observed from Figure 8. The green contour depicts the model predicted epithelial region and the blue contour represents the manually drawn epithelial ground truth. The pseudo-code for the proposed pipeline is presented in Algorithm 1.



Figure 7. Post-processing: (a) clean mask and (b) mask edge smoothing.

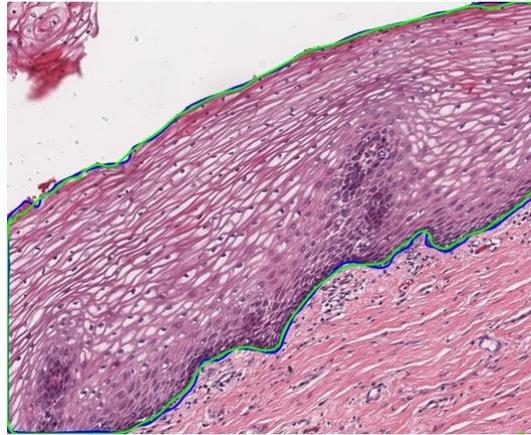


Figure 8. Segmentation contour.

---

Algorithm 1: Epithelial segmentation

---

```

% Preprocess
Generate  $(n, m)$  patches with stride  $s$ 
Calculate the respective ground-truth probabilities,

$$y_{gt}^k = \frac{1}{mn} \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} p_{mask}^k(x, y)$$

% Train
Initialize weights and bias
For  $i=1$ :  $N\_epochs$ , do
    Forward Pass, predict  $\hat{y}^k$ 
    L1 Loss:  $L = \sum_{k=1}^n |y_{gt}^k - \hat{y}^k|$ 
    Backpropagate,
    Update weights with ADADELTA optimizer:  $\theta_{i+1} = \theta_i + \Delta\theta_i$ 
End For
Save model and weights
% Test
Load model and weights
Pad image:  $pad_r = n_r s - rem(M, n_r s)$ ,  $pad_c = n_c s - rem(N, n_c s)$ 
Slice image to  $(p, q)$  sub-images,

$$n_t = \left\lceil \frac{MN}{pq} \right\rceil, n_r = \lceil \sqrt{n_t} \rceil, n_c = \left\lceil \frac{n_t}{\sqrt{n_t}} \right\rceil$$

Generate  $(n, m)$  patches with stride 4
Predict the probability of each pixel
Combine the predictions to form a gradient mask
Upscale the mask by factor of 4
% Post-process
Threshold the mask
Smooth the mask edges with quadratic Bezier curve,

$$B(t) = P_{i+1} + (1-t)^2(P_i - P_{i+1}) + t^2 P_{i+2}$$


```

---

### 3. EXPERIMENTS

We performed the following experiments on the epithelium set: high resolution cervical histology microscopy images (10x magnification). Since we have abundant image data, we do not incorporate leave-one-out methodology in our approach. The available 351 cervical histology images were divided into disjoint training and test sets. A set of 40 images, 10 from each class, were considered as training images, as previously described, with the remaining 311 images utilized for testing. The models were tested with various color spaces: RGB, LAB, HSV, and YCrCb, also with individual and combinations of color spaces. EpithNet models were observed to perform better with the normalized RGB images than other color spaces. The normalization is performed by dividing every pixel by the brightest pixel intensity of the image, and the images were split into smaller patches to create a large data set with a standard size. The images are split into overlapping patches forming our training set with 254,514 sub-images, and the test set is dependent on the epithelial image under test since each image has different dimensions, hence changing the number of sub-images obtained from the epithelial image.

#### 3.1. EXPERIMENTAL MODELS

We developed three models: EpithNet-16, EpithNet-32 and EpithNet-64. Each model is fed with different spatially localized images of sizes  $16 \times 16$ ,  $32 \times 32$  and  $64 \times 64$ , respectively. The models also vary in depth of six, seven and eight layers, respectively. We observed that the model with higher spatial information about a pixel's

surroundings has better knowledge with more feature information and can better predict the pixel's behavior. This is clearly evident from Table 2. As a step towards improving the segmentation accuracy, we have combined all three proposed models with internal interactions as shown in Figure 9 and named the model as EpithNet-mc (mc denotes multi-crop).

The EpithNet-mc model is designed to read an input image of size  $64 \times 64$  and at each layer, the input image at the first layer and the feature maps in the hidden layers are center cropped such that an array of size  $p \times q \times n$  is extracted from an array of size  $2p \times 2q \times n$ . These cropped versions of feature maps are concatenated with the feature maps from the lower resolution CNN model running in parallel. The dimensions of the concatenated feature maps are lowered by applying a  $1 \times 1$  convolutional filter. This reduces the dimensions of the feature maps while retaining salient features. The features coming from the higher resolution CNN have better feature information with additional knowledge of the spatial data, especially on the edges.

We compare EpithNet-16, EpithNet-32, EpithNet-64 and EpithNet-mc with UNet [11], which is a state-of-the-art transfer network model for image segmentation in the field of biomedical imaging. We modified UNet to make it capable of reading  $64 \times 64$  patch image data with the same set of sub-images subjected to training by EpithNet-64 to form our new baseline and named as UNet-64 with a structure containing 24 convolutional layers. UNet-64 is trained for 300 epochs with an Adam optimizer with a learning rate of 0.0001 under early stopping conditions. UNet is a fully convolutional neural network (FCN) and generates an output mask of size equal to the input image; the ground truths are the masks of the corresponding patch image data. In the testing phase,



Table 1. Complexity of baseline, UNet-64 and the proposed models.

Model	UNet-64	EpithNet-16	EpithNet-32	EpithNet-64	EpithNet-mc
Parameters ( $\times 10^6$ )	31.032	1.071	1.669	3.013	6.856

### 3.2. EXPERIMENTAL RESULTS

The image segmentation is evaluated using the following metrics:

$$J = \frac{|X \cap Y|}{|X \cup Y|} \quad (4)$$

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|} \quad (5)$$

$$PA = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (6)$$

$$MI = \sum_i \frac{n_{ii}}{(t_i + \sum_j n_{ji} - n_{ii})} \quad (7)$$

$$FWI = \frac{1}{\sum_i t_i} \sum_i \frac{t_i n_{ii}}{(t_i + \sum_j n_{ji} - n_{ii})} \quad (8)$$

where  $X$  and  $Y$  denote the binary masks of ground truth and predictions respectively; TP is the number of true positives denoting the pixels correctly identified as epithelium, TN is the number of true negatives that indicate the pixels correctly identified as background pixels, FP is the number of false positives indicating background pixels that are incorrectly identified as part of epithelium, FN is the number of false negatives indicating epithelium pixels mislabeled as background pixels;  $n_{ji}$  represents the number of pixels of class  $j$  predicted that actually belong to class  $i$  and  $t_i$  denotes the total number of pixels of class  $i$  in the ground truth mask.

The Equations (4)-(8) represent Jaccard index,  $J$ ; Dice score,  $DSC$ ; pixel accuracy,  $PA$ ; mean intersection over union (IOU),  $MI$ ; frequency weighted IOU,  $FWI$ ,

respectively. Jaccard index is defined as the number of pixels in the intersection of the two masks divided by union of pixels among the two masks. The Dice score is twice the number of common pixels divided by sum of pixel counts for both masks. Both Jaccard index and Dice score are the best descriptors of similarity coefficients between two masks and have been used in international segmentation challenges [18]. Pixel accuracy represents the percentage of pixels that were classified correctly. Mean IOU and frequency weighted IOU are measures of object detector accuracy. These metrics are more effective with multiclass segmentation problem settings even when the classes are imbalanced. *PA*, *MI* and *FWI* are the metrics considered for semantic segmentation and scene parsing [19]. All metrics penalize both false positive and false negative segmentation errors.

Table 2. Results on 311 cervical histology test data.

Model		<i>J</i>	<i>DSC</i>	<i>PA</i>	<i>MI</i>	<i>FWI</i>
UNet-64	median	0.738	0.849	0.845	0.709	0.740
	mean	0.676	0.789	0.822	0.692	0.712
	std	0.190	0.160	0.116	0.153	0.154
EpithNet-16	median	0.939	0.969	0.965	0.959	0.921
	mean	0.915	0.954	0.951	0.943	0.897
	std	0.070	0.043	0.045	0.049	0.081
EpithNet-32	median	0.947	0.973	0.970	0.966	0.933
	mean	0.931	0.964	0.961	0.954	0.916
	std	0.049	0.028	0.029	0.037	0.059
EpithNet-64	median	0.950	0.974	0.972	0.939	0.945
	mean	0.935	0.966	0.963	0.920	0.930
	std	0.049	0.028	0.032	0.062	0.054
EpithNet-mc	median	0.952	0.976	0.974	0.942	0.949
	mean	0.940	0.969	0.966	0.926	0.936
	std	0.041	0.023	0.026	0.052	0.046

The models, trained on patch image data generated from 40 original images, produced state-of-the-art segmentation results when tested on the 311 unseen image samples. The results indicate that the proposed models have performed better than the UNet-64 model on all the test images (see Table 2).

#### 4. DISCUSSION

We observe from Table 2 that the proposed EpithNet models outperform the baseline UNet-64 model. EpithNet-16, EpithNet-32 and EpithNet-64 are smaller CNN models with EpithNet-16 having 31 times fewer parameters than the UNet-64 model (Table 1). Higher resolution models like  $64 \times 64$  with EpithNet-64 have better segmentation results. This can be clearly understood from the fact that the model can have better awareness of its spatial environment with a higher resolution image which gives the ability to better judge the probability of the central pixel being an epithelium pixel. The multi-crop EpithNet (EpithNet-mc) model was found to improve the segmentation performance by 0.5% across all the metrics. The improvement is small but the intermediate connections from higher resolution CNN models to lower resolution CNN models help the combined model by providing better feature information across the lower resolution CNNs.

Our baseline model UNet-64, is found to learn features without overfitting, but during the testing phase, the model performed worse. This may be due to patch data which contain images where there is a complete epithelium region, complete background or the edge regions which contain both epithelium and background. Since the UNet

model is trained with a loss function that gives additional weight to the pixels at the border of the segmented objects, the images with complete epithelium or complete background confuse the model while learning features. The only advantage with the UNet-64 compared to EpithNet models is the shorter time taken to predict and generate a full binary mask. Despite having so many parameters UNet-64 is an FCN which generates a binary mask of size equal to the input image, whereas the EpithNet models predict the probability of individual pixels, which ultimately takes more time to predict the mask.

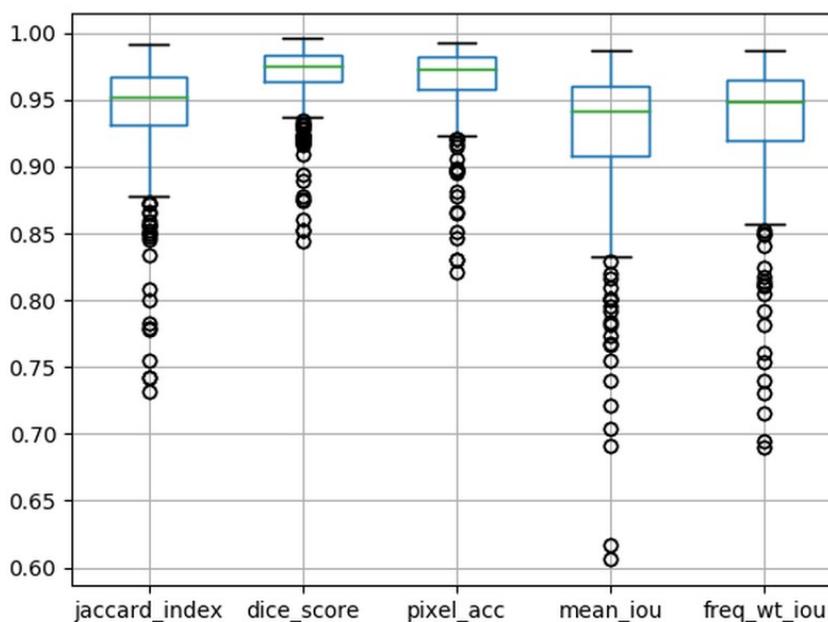


Figure 10. Boxplot of EpithNet-mc model with distribution of the metrics on 311 images.

The column parameters from left to right indicate Jaccard index, Dice score, pixel accuracy, mean intersection over union, frequency weighted intersection over union. See equations (4)-(9) above with accompanying parameter descriptions.

Due to limited histology data availability and problems with varying image sizes, the proposed EpithNet models are the best choice; this is clearly evident from the

segmentation results. The distribution of the metric values for test results on our best model (EpithNet-mc) can be visualized from Figure 10.

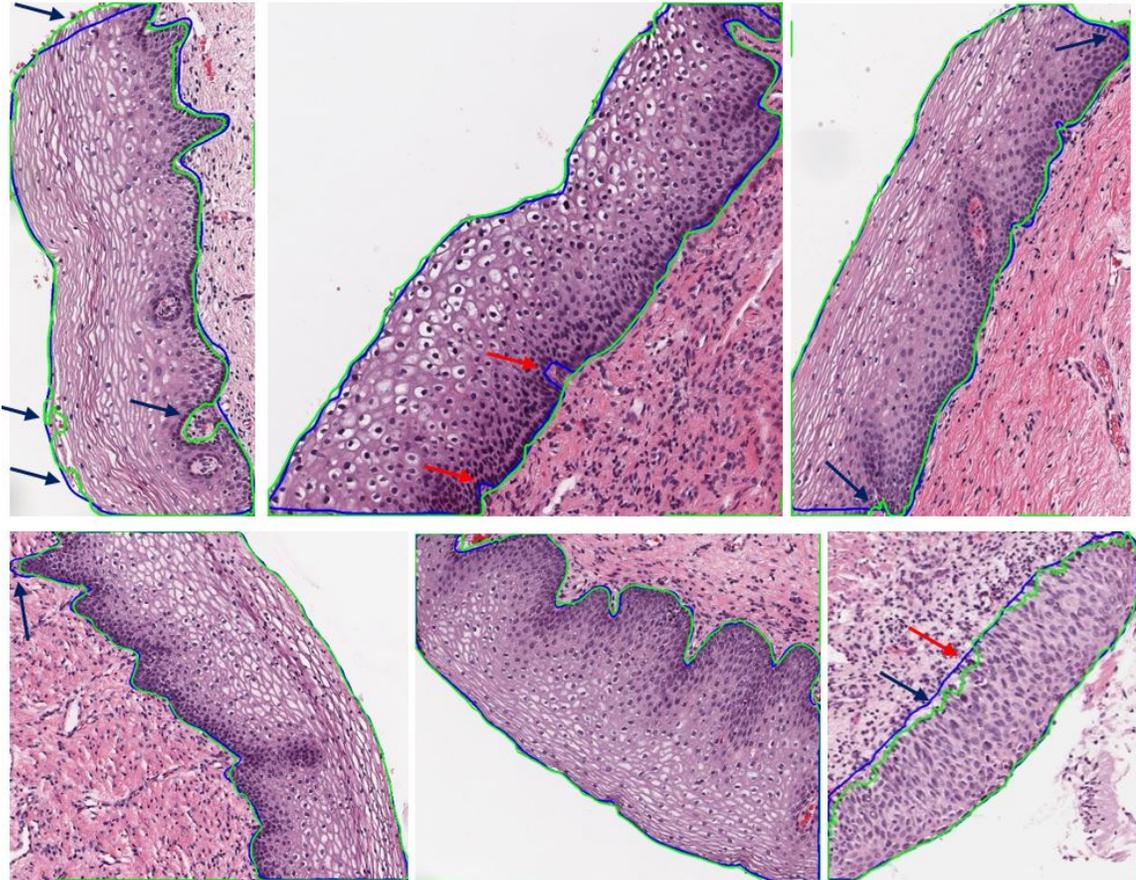


Figure 11. Segmentation results. Green contour represents the predicted mask and blue contour represent the ground truth mask. The blue arrows point to regions where the predicted masks do a better job in segmenting the epithelium regions compared to the manually drawn borders. The red arrows indicate regions of false segmentation.

Figure 11 shows some of the promising segmentation results of the epithelial regions from the test dataset. The segmentation at the edges fairly accurately tracks the ground truth edges. The model that generated the fuzzy mask is thresholded at 0.5 (empirically chosen as the optimum value), that is, pixel intensities greater than 0.5 are

considered as epithelial region and below 0.5 as background. It can also be observed from Figure 11 marked with blue arrows that sometimes the prediction masks do a better job in segmenting the epithelium accurately compared to the manually drawn ground truth masks. The automated segmentation tries to discard the regions that look similar to lamina propria or red blood cells near the edges and tries to include the epithelium regions, thereby correcting the manually drawn masks.

There were few exceptional cases that segmented a major portion of the epithelium region with small areas of false identification of the edges due to large variation in the staining and pattern of the nuclei looking similar to the nuclei in the stroma region below the epithelium. Sometimes the technique tries to remove the red blood cells even at the cost of missing epithelial regions. This can be observed from the red arrows in Figure 11.

## 5. CONCLUSION

We propose an approach to segment epithelial regions from a set of sparse epithelial data. Challenges in segmentation of histology images include variable staining, and noise including extravasated red blood cells and stain blobs, along with a limited number of ground truth images. The techniques here offer a deep-learning approach to meet the difficult challenge of architectural segmentation in automated histopathology. Reproducing the high-level approach of the expert pathologist is difficult. This article proposes a deep learning approach for architectural feature detection to replace handcrafted techniques that have been employed for such features [20].

The major contribution of this work is proposing a set of patch-based epithelium segmenting regression models yielding segmentation accuracy exceeding state-of-the-art results. We use a split-and-join scheme to optimally use the available memory during the testing phase and post-processing techniques to generate a smooth border using Bezier curves. The proposed EpithNet models are smaller and simpler but efficient in segmenting the epithelial regions of the cervical histology images. The generated mask is a probability mask, allowing the user to adjust the probability threshold to finely adjust the binary mask as needed. The results were reported by considering a default threshold value of 0.5. Moreover, it is observed from the results that the more the spatial information around a pixel is presented to the model, the better the segmentation masks generated, especially at the critical borders of the epithelium regions. EpithNet-mc was designed to combine the feature information from EpithNet-16, EpithNet-32 and EpithNet-64, which read image patches of varying spatial information centered at a given pixel. The features from layers carrying lower spatial dimensions were concatenated with features from layers carrying higher spatial dimensions to improve the quality of feature information, which ultimately resulted in generation of better segmentation masks. Although the baseline UNet-64 model, a fully convolutional network, is faster in generating the segmentation masks, the quality of the masks was poor. EpithNet, in contrast, not only generated relatively better epithelium masks, but also utilized fewer parameters, resulting in less GPU memory use.

The proposed models can also help in segmenting other epithelial tissues in pathology studies. Training these models with respective histology images would help in more accurate epithelium segmentation during the testing phase. The results of

segmentation of digital slides captured with different scanners and at varying resolutions is a subject for future research.

In future work, the proposed models will be used to generate the epithelium masks on digitized histology images at 10X magnification. These segmented regions will be further analyzed to ultimately create a classification model that can better estimate the severity of the cervical cancer by image processing. This could serve as a useful assistance tool for pathologists in segmenting out the useful regions and classifying the CIN levels while examining the samples.

## **ACKNOWLEDGEMENTS**

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

## **REFERENCES**

- [1] Global Cancer Observatory: Cancer Today.” World Health Organization: International Agency for Research on Cancer, 2018.  
<https://www.who.int/cancer/prevention/diagnosis-screening/cervical-cancer/en/>  
accessed August 21, 2019

- [2] B. A. Erickson, A. B. Olawaiye, A. Bermudez, Cervix Uteri, Chapter 52, in M. B. Amin, S. B. Edge, F. L. Greene, et al. (Eds.) *AJCC Cancer Staging Manual*. 8th Ed. New York: Springer; 2017.
- [3] J. Jeronimo, M. Schiffman, L. R. 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 *17th IEEE Symp. Computer-Based Medical Systems: IEEE Computer Society*, pp. 558-562, 2004.
- [4] V. Kumar, A. Abba, N. Fausto, and J. Aster, The Female Genital Tract, Chapter 22, in *Robbins and Cotran Pathologic Basis of Disease*, 9th ed., vol. 22, pp 1408. 2014.
- [5] L. He, L. R. Long, S. Antani, and G. R. Thoma, "Computer assisted diagnosis in histopathology in Sequence and Genome Analysis: Methods and Applications," *iConcept Press. Hong Kong*, vol. 15, pp. 271–287, 2011.
- [6] Y. Wang, D. Crookes, O. S. Eldin, S. Wang, P. Hamilton, and J. Diamond, "Assisted diagnosis of cervical intraepithelial neoplasia (CIN)," *IEEE J. Sel. Top. Signal Process.*, vol. 3, no. 1, pp. 112–121, 2009.
- [7] S. De *et al.*, "A fusion-based approach for uterine cervical cancer histology image classification," *Comput. Med. Imaging Graph.*, vol. 37, no. 7–8, pp. 475–87, 2013.
- [8] P. Guo *et al.*, "Nuclei-Based Features for Uterine Cervical Cancer Histology Image Analysis with Fusion-Based Classification," *IEEE J. Biomed. Heal. Informatics*, vol. 20, no. 6, pp. 1595–1607, 2016.
- [9] Y. Wang, D. Crookes, J. Diamond, P. Hamilton, and R. Turner, "Segmentation of Squamous Epithelium from Ultra-large Cervical Histological Virtual Slides," in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 775–778, 2007.
- [10] F. Wassmer, V. De Luca, R. Casanova, A. Soltermann, and G. Székely, "Automatic segmentation of lung carcinoma in histological images using a visual dictionary," in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pp. 1037–1040, 2016.
- [11] Z. Loncova, L. Hargas, and D. Koniar, "Segmentation of Microscopic Medical Images Using Local Binary Patterns Method," in *Proceedings of the 8th International Conference on Signal Processing Systems*, pp. 36–40, 2016.

- [12] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, pp. 234–241, 2015.
- [13] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv*, vol. *abs/1409.1556*, 2015.
- [14] V. Nair and G. E. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines," *Proc. 27th Int. Conf. Mach. Learn.*, no. 3, pp. 807–814, 2010.
- [15] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier Nonlinearities Improve Neural Network Acoustic Models," *Proc. 30th Int. Conf. Mach. Learn.*, vol. 28, p. 6, 2013.
- [16] M. D. Zeiler, "ADADELTA: An Adaptive Learning Rate Method," *arXiv*, vol. *abs/1212.5*, 2012.
- [17] M. Hazewinkel, "The Bezier Curve," in *Mathematics for Computer Graphics Applications*, 2nd ed., New York: Industrial Press Inc., 1999, p. 264.
- [18] N. C. F. Codella, D. Gutman, M. E. Celebi, "Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), Hosted by the International Skin Imaging Collaboration (ISIC), *ArXiv.org*>*cs*>*arXiv:1710.05006*, 2017. available at <https://arxiv.org/abs/1710.05006>
- [19] E. Shelhamer, J. Long, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, 2017.
- [20] J. Arevalo, A. Cruz-Roa, V. Arias, E. Romero, F. A. González FA, "An unsupervised feature learning framework for basal cell carcinoma image analysis," *Artif. Intell. Med.* vol 64, no. 2, pp. 131-45, 2015.

### **III. DEEPCIN: ATTENTION-BASED CERVICAL HISTOLOGY IMAGE CLASSIFICATION WITH SEQUENTIAL FEATURE MODELLING FOR PATHOLOGIST-LEVEL ACCURACY**

Sudhir Sornapudi<sup>1</sup>, R. Joe Stanley<sup>1</sup>, William V. Stoecker<sup>5</sup>, Rodney Long<sup>2</sup>, Zhiyun Xue<sup>2</sup>,  
Rosemary Zuna<sup>3</sup>, Shellaine R. Frazier<sup>4</sup>, Sameer Antani<sup>2</sup>

<sup>1</sup>Department of Electrical and Computer Engineering, Missouri University of Science and Technology, Rolla, MO

<sup>2</sup>Lister Hill National Center for Biomedical Communications for National Library of Medicine, National Institutes of Health, Bethesda, MD

<sup>3</sup>Department of Pathology, University of Oklahoma Health Sciences Center, Oklahoma City, OK

<sup>4</sup>Surgical Pathology Department, University of Missouri Hospitals and Clinics, Columbia, MO

<sup>5</sup>Stoecker & Associates, Rolla MO

#### **ABSTRACT**

Cervical cancer is one of the deadliest cancers affecting women globally. Cervical intraepithelial neoplasia (CIN) assessment using histopathological examination of cervical biopsy slides is subject to interobserver variability. Automated processing of digitized histopathology slides has the potential for more accurate classification for CIN grades from normal to increasing grades of pre-malignancy: CIN1, CIN2 and CIN3. Cervix disease is generally understood to progress from the bottom (basement membrane) to the top of the epithelium. To model this relationship of disease severity to spatial distribution of abnormalities, we propose a network pipeline, DeepCIN, to analyze high-resolution epithelium images (manually extracted from whole-slide images) hierarchically by focusing on localized vertical regions and fusing this local information for determining Normal/CIN classification. The pipeline contains two classifier networks:

1) a cross-sectional, vertical segment-level sequence generator is trained using weak supervision to generate feature sequences from the vertical segments to preserve the bottom-to-top feature relationships in the epithelium image data; 2) an attention-based fusion network image-level classifier predicting the final CIN grade by merging vertical segment sequences. The model produces the CIN classification results and also determines the vertical segment contributions to CIN grade prediction. Experiments show that DeepCIN achieves pathologist-level CIN classification accuracy.

**Index Terms:** Attention networks, cervical cancer, cervical intraepithelial neoplasia, classification, convolutional neural networks, digital pathology, histology, fusion based classification, recurrent neural networks.

## 1. INTRODUCTION

Cervical cancer prevention remains a big global challenge. It is estimated that in 2020 in the US 13,800 women will be diagnosed with invasive cervical cancer, and among them 4,290 will die [1]. This cancer ranks second in fatalities among 20-39 year old women [1]. Screening has helped decrease the incidence rate of cervical cancer by more than half since the mid-1970s through early detection of precancerous cells [2], yet 300,000 women die every year worldwide [3]. As a public health priority in 2018 the WHO director general made a global call for elimination of cervical cancer [4].

If clinically indicated, the cervix is further examined by taking a sample of cervical tissue (biopsy). The tissue sample is transferred to a glass slide and observed under magnification (histopathology). Cervical dysplasia or cervical intraepithelial

neoplasia (CIN) is the growth of abnormal cervical cells in the epithelium that can potentially lead to cervical cancer. CIN is usually graded on a 1-3 scale. CIN 1 (Grade I) is mild epithelial dysplasia, confined to the inner one third of the epithelium. CIN 2 (Grade II) is moderate dysplasia, usually spread within the inner two-thirds of the epithelium. CIN 3 (Grade 3) is carcinoma in-situ (severe dysplasia) involving the full thickness of the epithelium [5]. A diagnosis of Normal indicates the absence of CIN.

Figure 1 depicts the localized regions with all four classes.

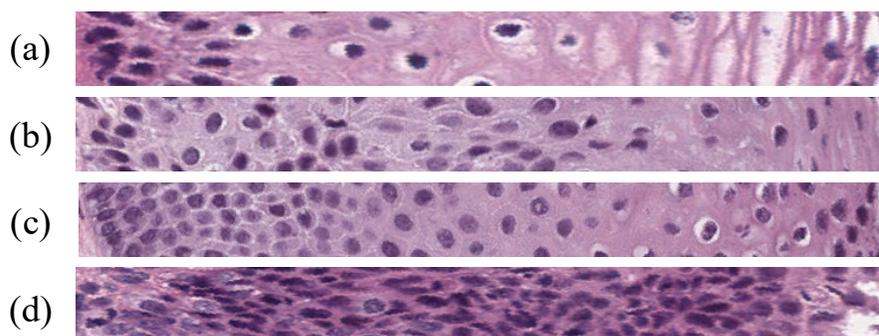


Figure 1. Sections of epithelium region with increasing CIN severity (from (b)-(d)) showing delayed maturation with increase in immature atypical cells from bottom-to-top. The sections can be categorized as (a) Normal, (b) CIN1, (c) CIN2, and (d) CIN3. In these images left-to-right corresponds to bottom-to-top of the epithelium.

Our previous work on computational approaches for digital pathology image analysis has relied mostly on extraction of handcrafted features based on the domain expert's knowledge. Guo et al. [6] manually extracted traditional nuclei features for CIN grade classification. The images were split into ten equal vertical segments for extraction of local features, and classified using voting fusion with support vector machine (SVM) and linear discriminant analysis (LDA). Huang et al. [7] used the LASSO algorithm for feature extraction with SVM ensemble learning for classification of cervical biopsy

images. Automated CIN grade diagnosis was also performed through analyzing Gabor texture features with K-means clustering [8] and slide level classification with texture features [9]. Accuracy fell short of that needed for clinical or laboratory use. In the past decade, success of deep learning approaches for image segmentation and classification in the health domain has attracted more research [10]. Toward that, AlMubarak et al. [11] developed a fusion-based hybrid deep learning approach that combined manually extracted features and convolutional neural network (CNN) features to detect the CIN grade from histology images. Li et al. [12] proposed a transfer learning framework with the Inception-v3 network for classifying cervical cancer images. An excellent review of computer vision approaches for cervical histopathology image analysis was presented in Li et al. [13].

A critical problem with manual CIN grading by pathologists is the variability among general pathologists in CIN determination. Stoler et al. [14] found agreement for the general community pathologist with the expert pathologist panel assignment to range from 38% to 68%: 38.2%, 38%, and 68% for CIN grades 1, 2 and 3, respectively. The overall Cohen's kappa value ( $\kappa$ ) was 0.46 for four grades, these three CIN grades and cervical carcinoma. Cai et al. [15] found close agreement among expert pathologists. For four expert pathologists, with 8-30 years of grading CIN slides, a weighted  $\kappa$  range of 0.799 to 0.887 was found. If automated CIN grading results can be made as close to expert readings as the variability among expert pathologist readings, automated CIN grading may become feasible.

Our proposed DeepCIN pipeline draws inspiration from the way pathologists examine epithelial regions under the microscope. They do not scan the entire slide at

once, instead they analyze local regions across the epithelium to understand the bottom-to-top growth of atypical cells and to compare the relative sizes of the cell nuclei in local neighborhoods. They use this local information to decide the CIN grade globally for the whole epithelium region. We developed a pathologist-inspired automated pipeline analogous to human study of histopathology slides, where we first localize the epithelial regions, then we analyze the features across these regions in both directions; finally, we fuse the feature information to predict the CIN class label and estimated the contribution of these local regions towards the global class result.

In this paper, we present DeepCIN, to automatically categorize high-resolution cervical histology images into Normal or one of the three CIN grades. Images used in this work are manually segmented epithelium regions extracted from digitized whole slide images (WSIs) at 10X magnification. The classification is carried out through hierarchical analysis of local epithelial regions by focusing on individual vertical segments and then combining the localized feature information in spatial context by introducing recurrent neural networks (RNNs).

The use of RNNs [16], [17] has been found to be successful in solving time-series and sequential prediction problems. Their use has led to better understanding of contextual features from images when combined with CNN-based models. Typically, CNNs act as a feature extractor, and RNNs learn the contextual information. Shi et al. [18] proposed a convolutional recurrent neural network for scene text (sequence-to-sequence) recognition. Attention mechanisms [19] were incorporated later to improve performance [20], [21]. Attention-based networks have been used in speech, natural language processing, statistical learning and computer vision [22].

A key aspect of our model is that it focuses on differentially informative vertical segment regions. This is crucial for deciding the level of CIN, because variation of CIN grade in local region could impact the overall CIN assessment of the epithelium [23]. The major contributions of this paper are:

- 1) Hierarchical image analysis from localized regions to the whole epithelium image.
- 2) Capturing the varying nuclei density across the epithelium region by vertically splitting the region into standard width segments with reference to the medial axis.
- 3) Weakly supervised training scheme for vertical segments.
- 4) Image-to-sequence two-stage encoder model for extracting localized segment level information.
- 5) Attention-based fusion (many-to-one model) for whole epithelium image CIN classification.
- 6) Identifying local segment contributions towards the whole image CIN classification.

## **2. METHODOLOGY**

DeepCIN incorporates a two-fold learning process (Figure 2). First, generated vertical segments from the epithelial image are fed to a two-stage encoder model for weak supervision training to constrain the segment class to the image class. Second, an attention-based fusion network is trained to learn the contextual feature information from

the sequence of segments and classify the epithelial image into one of the four classes.

The remainder of this section of the paper is organized as follows: Section 2.1 discusses cross-sectional vertical segment generation within an epithelium image; Section 2.2 and Section 2.3 present the two parts of the model: a segment-level sequence generator and an image-level classifier; Section 2.4 describes the model training approach.

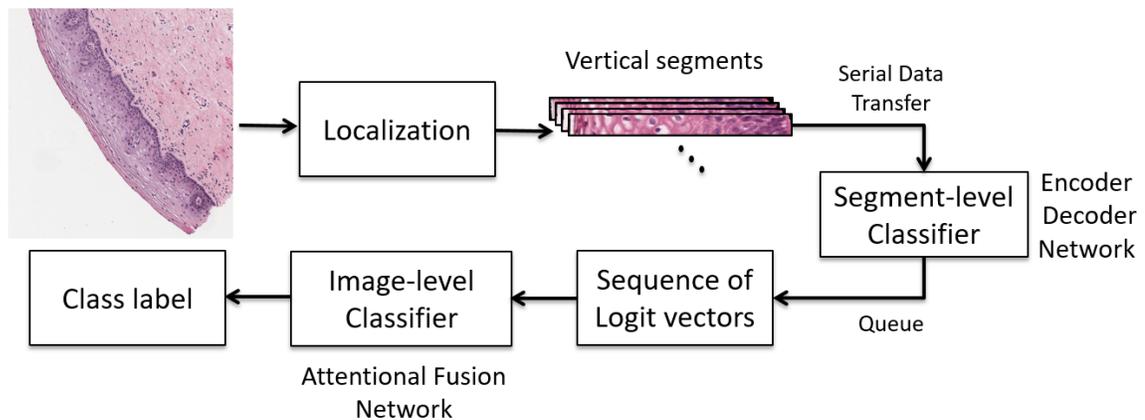


Figure 2. Overview of DeepCIN model.

## 2.1. LOCALIZATION

Initially, we process the manually segmented epithelium regions to find the medial axis and reorient the epithelium to be aligned horizontally, as performed by Guo et al. [6]. Guo’s methods are modified to generate standard-width vertical segments with reference to the medial axis. This helps in better understanding the pattern of atypical cells under uniform epithelium sections and generating more image data for training our deep learning model. We approximate the medial axis curve as a piece-wise linear curve by iteratively drawing a series of circles (left to right) of radii equal to the desired segment width. The center of each successive circle is the right-most intersection point of

the previously drawn circle and the medial axis curve. All the consecutive intersection points along the medial axis curve are joined to form a polygonal chain. At the midpoint of each line segment, we compute the slope corresponding to an intersecting perpendicular line. At the end points of the line segment, we draw vertical lines parallel to this midpoint perpendicular. This creates rectangular vertical regions of interest as shown in Figure 3. Using these individual vertical regions, we compute a bounding box, which we apply to the original image to crop a refined vertical segment. The heights and counts of vertical segments created in this manner vary with the shapes and sizes of the epithelial images. The height and width of the segments are empirically chosen to be 704 pixels and 64 pixels, respectively (for details refer to Section 3.1). The RGB image segments are further processed by channel-wise normalizing the pixel intensities with zero mean and standard deviation of value one, and rotating counter-clockwise by 90 degrees. This facilitates the classification of localized epithelial regions.

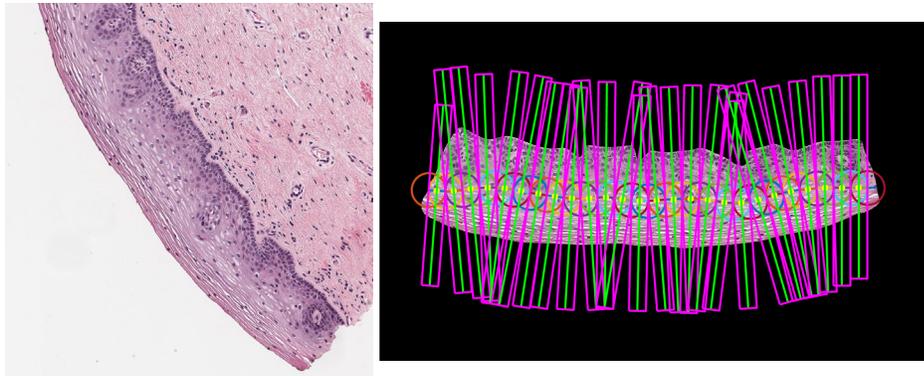


Figure 3. Localized vertical segment generation from an epithelial image.

Formally, we assume that an epithelial image  $I_{epth}$  has  $N$  vertical segments  $I_{vs_i}$  stacked up in a sequence by their spatial positioning from left-to-right such that:

$$I_{epth} = \{I_{vs_1}, I_{vs_2}, \dots, I_{vs_N}\}. \quad (1)$$

## 2.2. SEGMENT-LEVEL SEQUENCE GENERATION

The segment-level sequence generator network is built as a two-stage classifier model. The main objective of this network is to generate logit vectors to serve as localized sequence information for further image-level analysis. Because ground-truth labels for our vertical segments are not available, the network is trained against the image-level CIN grade. Since we expect variability in the true CIN grades across the vertical segments, use of the single image-level grade for all segments within an image introduces noisy labelling for the segments, and this may be expected to affect our training. Hence, we consider this a weakly supervised learning process.

We tackle this classification problem as a sequence recognition problem. As shown in Figure 4, the stage I encoder is constructed with a CNN that can extract the convolutional feature maps. These spatial features are then reduced to have height of 1 with maximum pooling operation. It is further transformed into a feature sequence by splitting along its width and concatenation of vectors formed by joining across the channels, similar to Shi et al. [18]. The RNN acts as a stage II encoder model that further encodes the sequential information to predict the class value (many-to-one model). It is important to understand that the vertical segments carry valuable localized feature information including varying nuclei density, which is crucial in the decision process. Therefore, it is well represented as a feature sequence and a bidirectional RNN focuses

on the intrinsic details within these vertical segment regions from left-to-right and right-to-left.

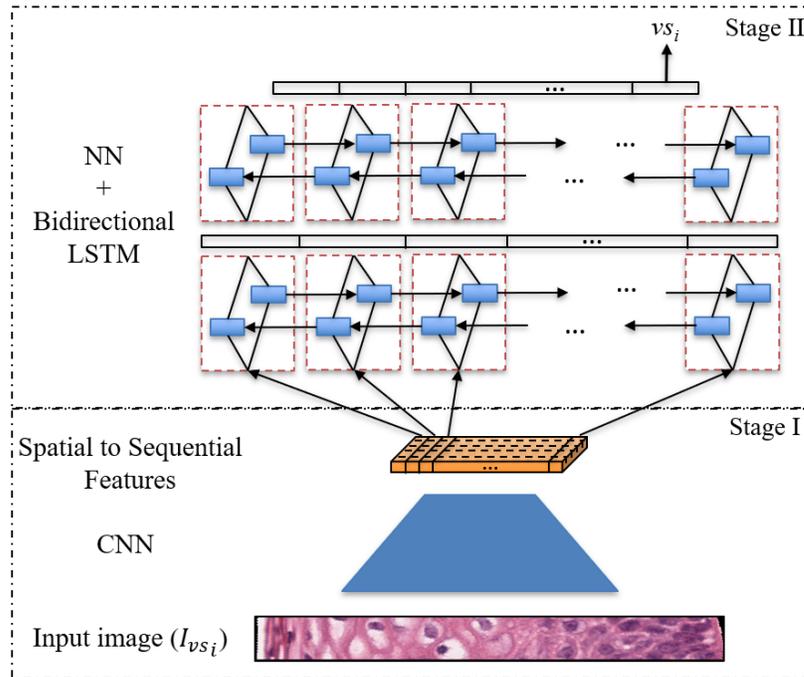


Figure 4. Segment-level sequence generator network with two-stage encoder structures.

The architecture of the proposed segment-level sequence generator is given in Table 1. The stage I encoder is built with first 87 layers of the DenseNet-121 model [24]. A max-pooling layer is added to this last layer such that the feature map has the height of 1. This can be considered as a feature sequence generated from left to right. Note that the convolutions always operate on local regions and hence are translationally invariant. Hence, the pixels in the feature maps from left-to-right corresponds to a local region in the original image (receptive field) from left-to-right. That is, the elements in the feature

sequence are image descriptors in the same order. Importantly, they preserve the bottom-to-top spatial relationships in the original epithelium image.

Table 1. Segment-level sequence generator model architecture.

	Layers	Configurations	Size
Stage I	Input	-	$3 \times 64 \times 704$
	Transition Layer 0	$k: 7 \times 7, s: 2, p: 3$	$64 \times 32 \times 352$
		$mp: 3 \times 3, s: 2, p: 1$	$64 \times 16 \times 176$
	Dense Block 1	$\begin{bmatrix} k: 1 \times 1, s: 1, p: 1 \\ k: 3 \times 3, s: 1, p: 1 \end{bmatrix} \times 6$	$256 \times 16 \times 176$
	Transition Layer 1	$\begin{bmatrix} k: 1 \times 1, s: 1 \\ ap: 2 \times 2, s: 2 \end{bmatrix}$	$128 \times 8 \times 88$
	Dense Block 2	$\begin{bmatrix} k: 1 \times 1, s: 1, p: 1 \\ k: 3 \times 3, s: 1, p: 1 \end{bmatrix} \times 12$	$512 \times 8 \times 88$
	Transition Layer 2	$\begin{bmatrix} k: 1 \times 1, s: 1 \\ ap: 2 \times 2, s: 2 \end{bmatrix}$	$256 \times 4 \times 44$
	Dense Block 3	$\begin{bmatrix} k: 1 \times 1, s: 1, p: 1 \\ k: 3 \times 3, s: 1, p: 1 \end{bmatrix} \times 24$	$1024 \times 4 \times 44$
Pooling	$mp: 4 \times 1, s: 1$	$1024 \times 1 \times 44$	
Stage II	BLSTM + NN	$nh: 256$	$512 \times 44$
		$nh: 256$	$256 \times 44$
	BLSTM + NN	$nh: 256$	$512 \times 44$
		$nh: 4$	$4 \times 44$
	Output	-	$4 \times 1$

$k$ ,  $s$ ,  $p$ ,  $mp$ ,  $ap$ , and  $nh$  are kernel, stride size, padding size, max pooling, average pooling and number of hidden layers, respectively. ‘BLSTM’ and ‘NN’ stands for bi-directional LSTM and single layer neural network, respectively.

To further analyze this feature context, the generated feature sequence is fed to a stage II encoder model built of RNNs. Specifically, we employed Bidirectional Long-Short-Term Memory (BLSTM) [25] networks to analyze and capture the long-term dependencies of the sequence from both directions. For the stage II encoder, two sets of

BLSTM and single layer neural networks (NN) were appended to the last max-pooling layer of the stage I encoder. The final classification result is extracted from the logit vector of the last element in the output sequence generated at the stage II encoder. These logit vectors summarize the information of all the vertical segments and, when combined, form an information sequence that is fused to determine the image-level CIN classification.

Assuming an epithelial image with  $N$  vertical segments  $I_{vs_i}$ , we have created logit sequence vectors  $vs_i$  obtained with a segment-level sequence generator  $f_s(\cdot; \theta)$ :

$$vs_i = f_s(I_{vs_i}; \theta) \quad (2)$$

where,  $\theta$  represents the model parameters.

### 2.3. IMAGE-LEVEL CLASSIFICATION

The image-level classifier network is designed as an attention-mechanism based fusion network as shown in Figure 5. We aim to capture the dependencies among vertical segments with a gated recurrent unit (GRU) [17]. The input sequences are picked up by a GRU, which tracks the state of the sequences with a gating mechanism. The output is a sequence vector that represents the image under test. We use a small classifier with an attentional weight for each GRU cell output to encode the sequence of the vertical segments as:

$$h_i = GRU(vs_i; h_{i-1}) \quad (3)$$

where  $i \in [1, N]$  and  $h_i$  is the hidden state that summarizes the information of the vertical segment  $I_{vs_i}$ .

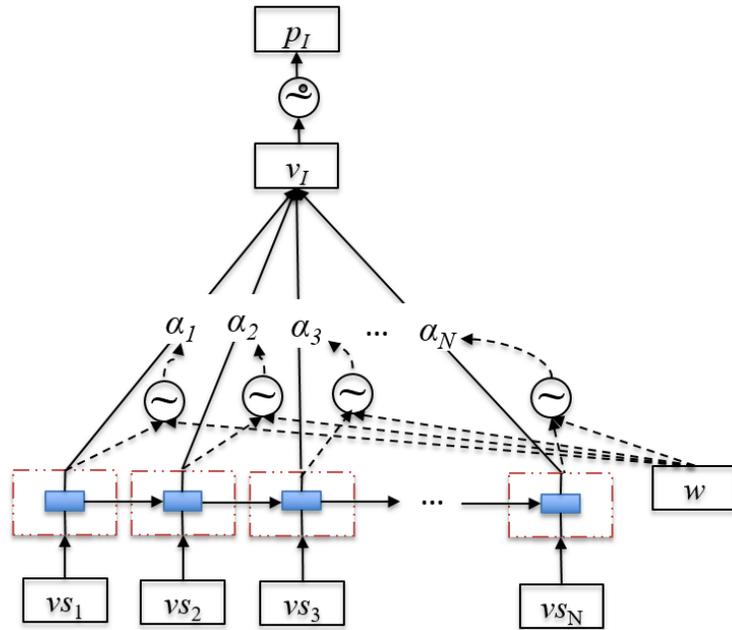


Figure 5. Attention-based fusion network for epithelial image-level classification. The input sequences are fed to GRU cells.  $\odot$  denote a two-layer neural network (NN) with hyperbolic tangent and Softmax activation functions, respectively to generate attentional weights.  $\odot$  denotes a single layer NN with Softmax activation function that produces the classification output.

The vertical segments may not contribute equally to epithelial image classification. We use an attention mechanism with a randomly initialized segment-level context vector  $w$  applied on the outputs of the GRU units that were subjected to tanh activated neural network. This vector is used to generate the attentional weights  $\alpha_i$  which analyze the contextual information and give a measure of importance of the vertical segments. The following equations explain the employed attention mechanism:

$$e_i = w^T \tanh(W_{vs} h_i + b_{vs}) \quad (4)$$

$$\alpha_i = \frac{\exp(e_i)}{\sum_{i=1}^N \exp(e_i)} \quad (5)$$

$$v_I = \sum_{i=1}^N \alpha_i h_i, \quad (6)$$

where  $W_{vs}$  and  $b_{vs}$  are trainable weights and bias.  $v_I$  is the image feature vector that summarizes all the information of vertical segments in an epithelial image. The image-level classification is determined by:

$$p_I = \text{softmax}(W_0 v_I + b_0). \quad (7)$$

## 2.4. TRAINING

We trained the proposed networks independently with stratified K-fold cross validation split at the image-level. First the segment-level sequence generator is trained to generate the logit vectors of all the segments and then concatenated to form a sequence to further train the image-level classifier.

During segment-level sequence generation, the problem of class imbalance is solved by up-sampling the vertical segment images with image augmentations: randomly flipping vertically and horizontally, rotating with a range of 180 to -180 degree angles, changing hue, saturation, value and contrast, and applying blur and noise. The objective is to minimize the cross-entropy loss (Equation 8) calculated directly from the vertical segment image and its restricted ground-truth label given by

$$L_k = -\sum_{vs} \log \left( \frac{\exp(y_k)}{\sum_j \exp(y_j)} \right) \quad (8)$$

where  $k$  is the class label of vertical segment image  $vs$  and  $y_k$  is the  $k^{\text{th}}$  label element value in the logit vector. We use ADADELTA [26] for optimization since it automatically adapts the learning rates based on the gradient updates. The initial learning rate set to 0.01.

For image-level classification, we use the weighted negative log likelihood of correct labels to compute the cost function and back propagate the error to update the

weights with a stochastic gradient descent (SGD) optimizer (learning rate was fixed at 0.0001). Training loss is given by:

$$L'_k = -q_k \sum_I \log(p_{I_k}) \quad (9)$$

where  $k$  is the class label of epithelial image  $I$  and  $q_k$  is the weight of the label  $k$ .

### 3. EXPERIMENTS

We conducted experiments on our cervical histopathology image database to evaluate the effectiveness of the proposed classification model and compared its performance with other state-of-the-art methods.

#### 3.1. DATASET AND EVALUATION METRICS

For all the cross-validation experiments, we use a dataset that contains 453 high-resolution cervical epithelial images extracted from 146 hematoxylin and eosin stained cervical histology WSIs. In addition, we use independent 224 high-resolution epithelium images as a hold-out test data. These WSIs were provided by Department of Pathology at the University of Oklahoma Medical Center in collaboration with the National Library of Medicine. They were scanned at 20X using Aperio ScanScope slide scanner and saved with the file extension sv. All images have corresponding ground-truth labels. These annotations were carried out by an expert pathologist. The epithelial images have varying sizes which range from about  $550 \times 680$  pixels (smallest) to  $7500 \times 1500$  pixels (largest). This varying size affects the number of vertical segments generated from an image, typically ranging from 6 to 118. Though the vertical segments are generated such that the

widths are 64 pixels wide, the height of these segments range from 160 to 1400 pixels. We address this problem by resizing the images to their median height: 704 pixels. This height was chosen empirically as a multiple of 32 in order to apply convolutions for feature extraction.

Table 2. Class label distribution from 453 epithelial images.

Class	Epithelial Images		Segments	
	Count	%	Count	%
Normal	244	53.8	6,836	57.7
CIN1	57	12.6	1,433	12.1
CIN2	79	17.5	2,039	17.2
CIN3	73	16.1	1,546	13.0
Total	453	100.0	11,854	100.0

The segments were pre-processed such that they are RGB images of standard size:  $64 \times 704 \times 3$ . We have created a total of 11,854 vertical segment images from 453 epithelial images. The class distribution of these data is shown in Table 2. There are two main challenges with this epithelial image dataset. First, the cervical tissues have irregular epithelium regions, with color variations, intensity variations, red stain blobs, variations in nuclei shapes and sizes, and noise and blurring effects created during image acquisition. These effects tend to have large inter- and intra-class variability across the four classes we seek to label. Second, even though our database is labelled by experts and may be considered of high quality, it is relatively small. This is a common and recognized problem in the biomedical image processing domain.

The scoring metrics used for the performance evaluation are precision (P), recall (R), F1-score (F1), classification accuracy (ACC), area under Receiver Operating

Characteristic curve (AUC), average precision (AP) and Matthews correlation coefficient (MCC). Cohen’s kappa score ( $\kappa$ ) is used for the evaluation of the scoring schemes described in Section 3.4. The percentage weighted average scores were reported due to the inevitable imbalance in the data distribution.

### 3.2. IMPLEMENTATION DETAILS

Although the entire DeepCIN model can be implemented end-to-end, we have split the process into two independent training steps. This model was chosen to overcome the GPU memory limitation to process these large input images and network architectures.

Details about the segment-level sequence generator network and image-level classifier network are given in Table 1 and Figure 4, respectively. Both networks output four classes. The first network is trained with weak supervision to determine the logit sequence vectors of each vertical segment. The class outputs of the final network comprise our major concern.

A transfer learning technique was incorporated in the stage I encoder of the segment-level sequence generator. The convolution filters were initialized with ImageNet [27] pre-trained weights and were left frozen since the stage I encoder is built with initial layers of the DenseNet-121 model which presumably has weights already set to extract low-level image features such as edges, colors and curves. All the CNN layers are activated with the rectified linear unit (ReLU) function, and the single layer neural network followed by BLSTM layers in the stage II encoder, which does not impose any non-linearity to get logit vector sequence. The latter network consists of GRU cells (with

128 hidden units), a two-layer neural network (NN) with hyperbolic tangent and Softmax activation functions, respectively, to generate attentional weights, and a single-layer NN with Softmax activation function to produce the classification output from the image feature vector.

We trained and validated the models using stratified 5-fold cross validation. We split training and validation data at the image level and maintained the same distribution across both the models. To address the class imbalance problem, we have up-scaled the less populated class images with image augmentations for the segment level sequence generation and in the image level classification we employed a weighted loss function.\

Each individual fold for both the models were trained for 200 epochs with batch size of 56 with early stopping to avoid overfitting.

We implemented our localized vertical segment generation in MATLAB [28] running on an Intel Xeon CPU @ 2.10GHz which took 3.42 seconds on average to process one epithelial image. The deep learning models are trained under CUDA 10.2 and CuDNN v7.6 backend on an NVIDIA Quadro P4000 8GB GPU and 64GB RAM with a PyTorch v1.4 [29] framework. The time taken for validation is about 0.68 seconds per epithelial image. Thus, the entire DeepCIN pipeline takes 4.10 seconds on average to process and validate one epithelial image.

### **3.3. ABLATION STUDIES**

In this section, we perform classifier ablation studies on the DeepCIN pipeline to understand its key aspects. The experiments include comparison with different segment

widths, stage I and stage II encoder variants, different fusion techniques and benchmark models.

The proposed model takes standard size image inputs. Resizing images will cause image distortions. We observe that this has a minor effect on the performance, expected since both the training and testing images are similarly resized which would result in the model’s capability of handling such distortions. But the segment width is to some extent a free variable whose setting may modulate the amount of local spatial information contained in a vertical segment. Recognizing this, we experimented with segment widths of 32, 64 and 128. According to Table 3, we observe that a segment width of 64 pixels is an optimal choice (in our experimental search space) compared to the segments with 32 pixels wide and 128 pixels wide.

Table 3. Ablation study on segment widths.

Segment width	P	R	F1	ACC	AUC	AP	MCC
32	82.9	82.3	81.2	82.3	93.5	85.3	72.3
64*	88.6	88.5	88.0	88.5	96.5	91.5	82.0
128	85.3	85.6	84.9	85.6	95.9	89.8	77.1

The stage I encoder in the segment level sequence generator acts like a spatial feature extractor. Because our biomedical digital image environment is not data-rich for training deep learning models, we have experimented with various published models which have been pre-trained with the benchmark ImageNet database. Only a set of initial layers that extract low level features from the input image are considered in building the

stage I encoder. The top performing stage I encoder model results were recorded as shown in Table 4. We observed that DenseNet-121 was better at extracting the crucial epithelial information, compared to ResNet-101 [30] and Inception\_v3 [31] models. The DenseNet-121 model is better at feature reuse and feature propagation throughout the network with reduced parameters. Both DenseNet-121 and ResNet-101 are good at alleviating vanishing gradient problems, however DenseNet-121 with its feed-forward interconnections among layers helps in better feature understanding. Inception-v3 uses models that are wider rather than deeper to prevent overfitting with factorizing convolutions to reduce the number of parameters without compromising network efficiency.

Table 4. Ablation study on stage I encoder models.

Stage I Encoder	P	R	F1	ACC	AUC	AP	MCC
DenseNet-121*	88.6	88.5	88.0	88.5	96.5	91.5	82.0
ResNet-101	87.1	86.9	86.4	86.9	95.0	88.9	79.6
Inception-v3	85.5	85.4	85.1	85.4	94.8	87.8	77.1

The stage II encoder further encodes the feature sequence that is mapped from the translationally invariant feature information available from the stage I. Our efforts to use bidirectional LSTM in stage II encoder delivered better performance on the segment-level sequence generation that reflects on generating essential and better logit feature vectors. Table 5 shows that bidirectional analysis is enables understanding of the context of the feature information; this aided in up-sampling the segment data by flipping the input images horizontally. The use of attention was not helpful for understanding the feature sequence in the vertical segments with almost 1% decrease in performance across

all the metrics (Table 5). This indicates that the entire feature sequence is equally important to interpret the localized information, as shown by the equal distribution of attentional weights. The use of vanilla neural networks (fully connected layers) was comparatively less efficient because LSTMs contain internal state cells that act as long-term and short-term memory units and manage to learn by remembering the important information and forgetting the unwanted. Neural networks lack this ability and focus only on the very last input.

Table 5. Ablation study on stage II encoder models.

Stage II Encoder	P	R	F1	ACC	AUC	AP	MCC
BLSTM*	88.6	88.5	88.0	88.5	96.5	91.5	82.0
BLSTM + Attention	87.9	87.6	87.7	87.6	95.2	88.9	80.1
FC	85.3	85.0	84.2	85.0	94.7	87.4	76.3

We observed that attentional weights help analyze the valuable information from the contribution of each segment towards the image-level classification. Table 6 confirms this observation, showing nearly a 2% improvement in performance with inclusion of attention. Techniques like maximum voting and average voting of segment-level sequence generation results are simple and straight-forward, but fail provide the additional information about the localized segment data.

Table 6. Ablation study on fusion techniques.

Fusion	P	R	F1	ACC	AUC	AP	MCC
GRU	86.3	86.1	85.6	86.1	96.3	90.4	78.0
GRU+Attention*	88.6	88.5	88.0	88.5	96.5	91.5	82.0
Max vote	87.6	87.2	87.0	87.2	-	-	79.9
Avg vote	88.0	87.6	87.4	87.6	-	-	80.6

#### 4. PERFORMANCE OF DEEPCIN

We finally compare the performance of the proposed model with the state-of-the-art CIN classification models. The models used for the comparison are proposed by Guo et al. [6] and AlMubarak et al. [11]. The best model of Guo et al. [6], linear discriminant analysis, was trained with 27 handcrafted features extracted from vertical image segments. The epithelium was split into 10 equal parts to create these segments and fusion was performed through a voting scheme. AlMubarak et al. [11] used the same vertical segments and divided them into three sections: top, middle and bottom.  $64 \times 64$  size Lab color space image patches were extracted to train three CNN models. The resulting confidence values from these sections were treated as features, and the 27 features were concatenated to form a hybrid approach for training an SVM classifier. The final classifiers of both these models were trained with a leave-one-out approach.

Table 7. Comparison with state-of-the-art models.

Model	P	R	F1	ACC	AUC	AP	MCC
Guo <i>et al.</i> [6]	67.5	73.3	69.4	73.4	-	-	56.5
AlMubarak <i>et al.</i> [11]	66.1	75.6	70.4	75.5	90.9	78.1	60.3
Ours*	88.6	88.5	88.0	88.5	96.5	91.5	82.0

For a direct comparison, we have retrained [6] and [11] models on the 453 high-resolution epithelial histopathology image data. Table 7 shows that the proposed model performs best for the CIN classification task. Additionally, our model provides the significance of individual local regions towards the whole image classification. Results for sample images from the proposed DeepCIN model are shown in Figure 6. We

observed that the performance was uniform among different sizes of epithelium images. The distribution of the entire data and the predictions for all 5-folds is depicted in the Sankey diagram in Figure 7, which shows the proportion of images that are correctly classified and misclassified. Image samples belonging to the CIN1 class were mostly misclassified as normal class. Two reasons may explain this: 1) CIN1 images closely resemble normal images; 2) the number of CIN1 class images is small, relative to the number of Normal class images.

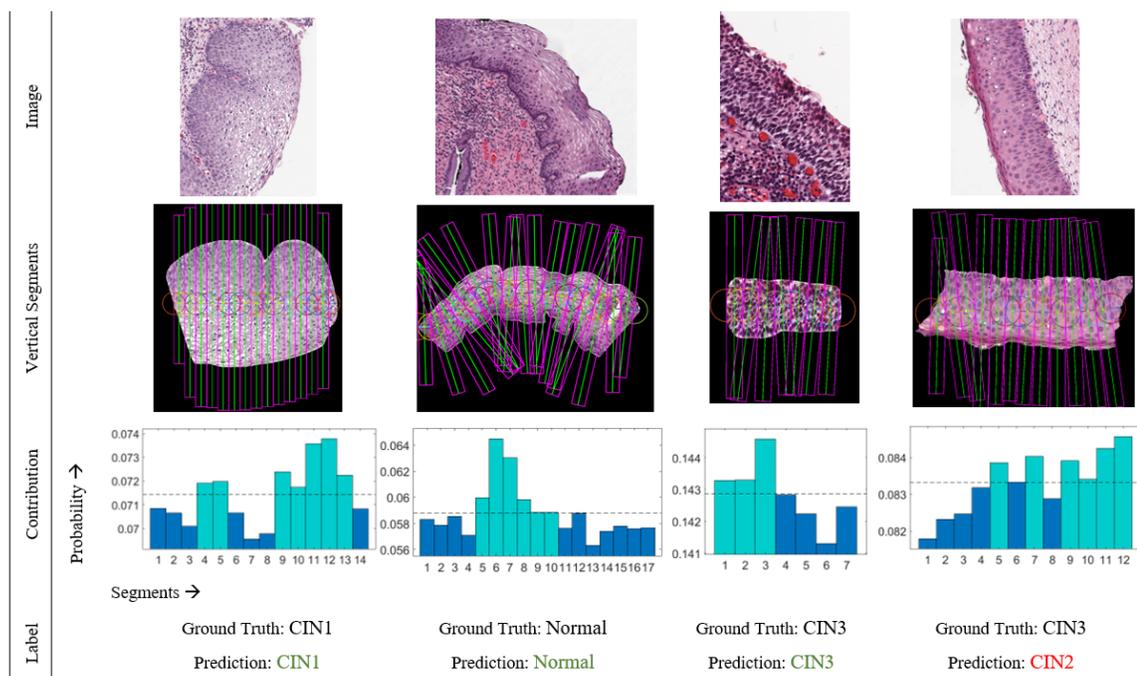


Figure 6. Results of DeepCIN. From top to bottom, each column presents original image, localized vertical regions, contribution of segments within an image towards the image-level CIN classification (represented as probability distribution over the segments (attentional weights), the dotted lines indicate mean value and segments above the mean value, highlighted in green, are contributing the most), and corresponding ground truth and prediction labels, respectively.

As an extension, we have tabulated the performance model with exact class labels, CIN versus Normal, CIN3-CIN2 versus CIN1-Normal, CIN3 versus CIN2-CIN1-Normal, and off-by-one class (Table 8). For the exact class label scheme, the predicted class label should exactly match the expert ground-truth class label. The CIN versus Normal scheme is an abnormal-normal grouping of the predicted labels. The CIN3-CIN2 versus CIN1-Normal and CIN3 versus CIN2-CIN1-Normal interclass grouping schemes resemble the clinical decisions for treatment. The Off-by-one scheme emphasizes the possible disagreement between the expert pathologists while labelling the CIN class which is usually observed to be one grade off [32].

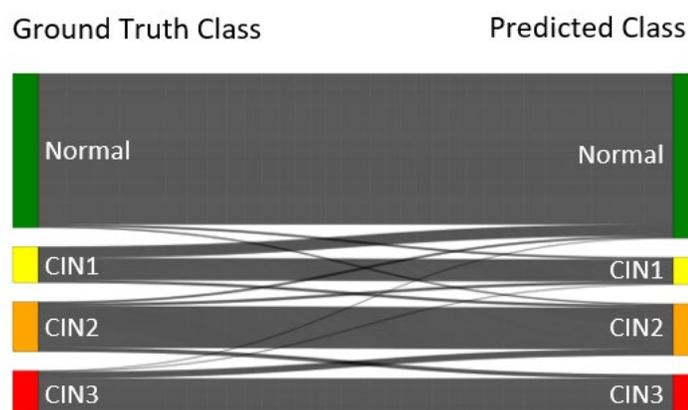


Figure 7. Sankey diagram – based on the combined test results from the 5-fold cross-validation. The height of each bar is proportional to the number of samples corresponding to each class.

We have ensembled our five models from the 5-fold cross validation with maximum voting system to test the model performance on unseen data. The results from the hold-out 224 image data are shown in Table 9. The results when compared with Table 8 indicate that the proposed model is good at generalizing on unseen data. We noticed

that the kappa score with CIN3 versus CIN2-CIN1-Normal scoring scheme is affected due to small portion of CIN 3 images were miss predicted as CIN 2 class.

Table 8. 5-fold cross-validation results with different scoring schemes.

Scoring Scheme	P	R	F1	ACC	AUC	AP	MCC	$\kappa$
Exact class label	88.6	88.5	88.0	88.5	96.5	91.5	82.0	81.5
CIN vs Normal	94.6	94.1	94.0	94.1	93.8	97.7	88.5	87.9
CIN3-CIN2 vs CIN1-Normal	96.8	96.7	96.7	96.7	96.0	98.9	92.7	92.5
CIN3 vs CIN2-CIN1-Normal	96.2	96.0	96.0	96.0	88.4	98.3	85.3	84.8
Off-by-one	-	-	-	98.9	-	-	-	-

Table 9. CIN classification results on 224 image-set.

Scoring Scheme	P	R	F1	ACC	AUC	AP	MCC	$\kappa$
Exact class label	90.2	88.4	88.2	88.4	98.0	93.1	80.5	80.0
CIN vs Normal	97.3	97.3	97.3	97.3	97.2	99.7	94.4	94.4
CIN3-CIN2 vs CIN1-Normal	95.7	95.6	95.5	95.5	94.0	99.1	90.3	90.0
CIN3 vs CIN2-CIN1-Normal	93.0	92.4	91.5	92.4	78.2	97.0	71.9	68.1
Off-by-one	-	-	-	98.2	-	-	-	-

## 5. DISCUSSION

The main objective of the DeepCIN model is to classify the high-resolution epithelium images into normal or precancerous transformation of cells of the uterine cervix. We generate classification results by fusing localized information, forming a sequence of logit feature vectors in the same order of the vertical segments from the

epithelium image. The number of vertical segments created varies since the epithelium images have arbitrary shapes. Traditional neural networks are limited to fixed-length input, but RNNs have the capability to read varying input sequences along with memorization. We employ a GRU to read the arbitrarily shaped input sequences. GRU with attention helps in better understanding the differentially informative localized data. Unlike the stage II encoder from the segment-level sequence generator, incorporation of attention helped the model to better fuse the segment data and identify localized regions that are significantly important in the classifying the epithelial image.

It is now four decades since Marsden Scott Blois presented a paradigm for medical information science to distinguish domains in medicine in which humans are essential from those in which computation is essential and computers are likely to play a primary role [33]. He emphasized the importance of human judgment in the former domain, which includes most of clinical medicine, but does not include the evaluation and interpretation of physiological parameters, for example blood gases, which is the proper domain of computers. With regard to the Blois paradigm, we propose that computer processing of histopathology images falls within the computational domain, and computers are likely to play a primary role.

## 6. CONCLUSION

In this study, we address the CIN classification problem by focusing on localized epithelium regions. The varying atypical nuclei density which is crucial in CIN determination is better analyzed by sequence mapping of the deep learning features. This

sequence is interpreted in both directions under weak supervision with the long-term and short-term memory of the feature information. We employed an attention-based fusion approach to carry out an image-level classification. This hierarchical approach not only produces the image-level CIN classification labels but also provides the contribution of each individual vertical segment of the epithelium towards the whole image classification. We conjecture that this information highlights the highest-risk areas; this serves as an automated check for the pathologist's assessment.

We observed that our proposed model, DeepCIN, has out-performed state-of-the-art models in classification accuracy. The final image-level classification accuracies and Cohen's kappa score are {88.5% ( $\pm$  2.2%), 81.5%}, {94.1% ( $\pm$  2.0%), 87.9%}, {96.7% ( $\pm$ 1.6%), 92.5%}, {96.0% ( $\pm$ 1.7%), 84.8%}, and {98.9% ( $\pm$  0.0%), -} for exact class label, CIN versus Normal, CIN3-CIN2 versus CIN1-Normal, CIN3 versus CIN2-CIN1-Normal and leave-one-out schemes, respectively. These results significantly exceed the variability of community pathologists when measured against the gold standard, and are in the range of inter-pathologist variability for expert pathologists as measured by the  $\kappa$  statistics.

Limitations of this work include use of a database that is not publicly available, which precludes validation by other researchers. Ground truth for the entire set was based on only one expert pathologist. Part of the set was scored by two pathologists; accuracies obtained for the two sets are similar.

Future work could improve results by including more annotated image data with balanced class distribution for training. There is also a possibility for improvements if the entire model could be trained end-to-end, which requires greater GPU resources. Our

future research will focus on WSI-level classification with end-to-end automation which combines the proposed model with our previous work on automated epithelium segmentation [34], and automated nuclei detection [35] for extracting enhanced feature information.

### ACKNOWLEDGMENTS

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

### REFERENCES

- [1] American Cancer Society. Cancer Facts & Figures 2020. Atlanta, GA 2020.
- [2] Islami F, Fedewa SA, and Jemal A. Trends in cervical cancer incidence rates by age, race/ethnicity, histological subtype, and stage at diagnosis in the United States. *Prev Med.* Baltimore; 2019. 123:316–323.
- [3] World Health Organization. Sexual and reproductive health, Prevention and Control of Cervical cancer. 2019. Available from: <https://www.who.int/reproductivehealth/topics/cancers/en/>. [Last accessed: 2020 Jun 09].
- [4] World Health Organization. Sexual and reproductive health, WHO Director-General calls for all countries to take action to help end the suffering caused by cervical cancer. 2018. Available: <https://www.who.int/reproductivehealth/call-to-action-elimination-cervical-cancer/en/>. [Last accessed: 2020 Jun 09].
- [5] Melnikow J, Nuovo J, Willan AR, B. Chan KS, and Howell LP. Natural History of Cervical Squamous Intraepithelial Lesions: A Meta-Analysis. *Obstet Gynecol* 1998; 92(4).

- [6] Guo P, Banerjee K, Joe Stanley R, Long R, Antani S, Thoma G, *et al.* Nuclei-based features for uterine cervical cancer histology image analysis with fusion-based classification. *IEEE J Biomed Health Inform* 2016; 20:1595-607.
- [7] Huang P, Zhang S, Li M, Wang J, Wang B, and Lv X. Classification of Cervical Biopsy Images Based on LASSO and EL- SVM. 2020.
- [8] Rahmadwati, Naghdy G, Ros M, Todd C and Norahmawati E. Cervical cancer classification using Gabor filters. *Proc. - 2011 1st IEEE Int. Conf. Healthc. Informatics, Imaging Syst. Biol. HISB 2011*; pp. 48–52.
- [9] Wang Y, Crookes D, Eldin OS, Wang S, Hamilton P and Diamond J. Assisted diagnosis of cervical intraepithelial neoplasia (CIN). *IEEE J Sel Top Signal Process* 2009;3(1):112–121.
- [10] Lecun Y, Bengio Y and Hinton G. Deep learning. *Nature* 2015; 521(7553):436–444.
- [11] AlMubarak H, Stanley J, Guo P, Long R, Antani S, Thoma G, *et al.* A Hybrid Deep Learning and Handcrafted Feature Approach for Cervical Cancer Digital Histology Image Classification. *Int J Healthc Inf Syst Informatics* 2019;14(2):66–87.
- [12] Li C, Xue, D, Zhou X, Zhang J, Zhang H, Yao Y, *et al.* Transfer learning based classification of cervical cancer immunohistochemistry images. *ACM Int Conf Proceeding Ser* 2019; pp. 102–106.
- [13] Li C, Chen H, Li X, Xu N, Hu Z, Xue D, *et al.* A review for cervical histopathology image analysis using machine vision approaches. *Artificial Intelligence Review* 2020.
- [14] Stoler MH, Ronnett BM, Joste NE, Hunt WC, Cuzick J, and Wheeler CM. The Interpretive Variability of Cervical Biopsies and Its Relationship to HPV Status. *In The American journal of surgical pathology* 2015.
- [15] Cai B, Ronnett B M, Stoler M, Ferenczy A, Kurman R J, Sadow D, *et al.* Longitudinal Evaluation of Interobserver and Intraobserver Agreement of Cervical Intraepithelial Neoplasia Diagnosis Among an Experienced Panel of Gynecologic Pathologists. *Am J Surg Pathol* 2007; 31(12).
- [16] Hochreiter S and Schmidhuber J. Long Short-Term Memory. *Neural Comput.* 1997; 9(8): 1735–1780.
- [17] Cho K, van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, *et al.* Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*; 2014; 1724–1734.

- [18] Shi B, Bai X, and Yao C. An End-to-End Trainable Neural Network for Image-Based Sequence Recognition and Its Application to Scene Text Recognition. *IEEE Trans Pattern Anal Mach Intell* 2017; 39(11):2298–2304.
- [19] Bahdanau D, Cho K, and Bengio Y. Neural Machine Translation by Jointly Learning to Align and Translate. *CoRR* 2014; 1409.
- [20] Shi B, Yang M, Wang X, Lyu P, Yao C, and Bai X. ASTER: An Attentional Scene Text Recognizer with Flexible Rectification. *IEEE Trans Pattern Anal Mach Intell* 2018;41(9):2035–2048.
- [21] Luo C, Jin L, and Sun Z. MORAN: A Multi-Object Rectified Attention Network for scene text recognition. *Pattern Recognit* 2019; 90:109–118.
- [22] Chaudhari S, Polatkan G, Ramanath R, and Mithal V. An Attentive Survey of Attention Models. 2019.
- [23] Wentzensen N, Walker J L, Gold M A, Smith K M, Zuna R E, Mathews C, *et al.* Multiple Biopsies and Detection of Cervical Cancer Precursors at Colposcopy. *J Clin Oncol* 2015; 33(1):83–89.
- [24] Huang G, Liu Z, Maaten L, and Weinberger K. Densely Connected Convolutional Networks. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2017; 2261–2269.
- [25] Graves A, Liwicki M, Fernández S, Bertolami R, Bunke H, and Schmidhuber J. A Novel Connectionist System for Unconstrained Handwriting Recognition. *IEEE Trans Pattern Anal Mach Intell* 2009; 31(5): 855–868.
- [26] Zeiler M D. ADADELTA: An Adaptive Learning Rate Method. *arXiv* 2012; 1212(5).
- [27] Deng J, Dong W, Socher R, Li L, Li K, and Fei-Fei L. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09* 2009.
- [28] MATLAB 9.7.0.1190202 (R2019b). The MathWorks Inc. Natick, Massachusetts; 2018.
- [29] Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, *et al.* Automatic differentiation in PyTorch. 2017.
- [30] He K, Zhang X, Ren S, and Sun J. Deep Residual Learning for Image Recognition. 2015.
- [31] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, and Wojna Z. Rethinking the Inception Architecture for Computer Vision. 2015.

- [32] Guo P. Data Fusion Techniques for Biomedical Informatics and Clinical Decision Support. ProQuest Diss. Theses 2018; p. 135.
- [33] Blois M S. Clinical Judgment and Computers. *N Engl J Med* 1980; 303(4): 192–197.
- [34] Sornapudi S, Hagerty J, Stanley RJ, Stoecker WV, Long R, Antani S, Thoma G, Zuna R, Frazier SR. EpithNet: Deep regression for epithelium segmentation in cervical histology images. *J Pathol Inform* 2020; 11:10. DOI: 10.4103/jpi.jpi\_53\_19.
- [35] Sornapudi S, Stanley RJ, Stoecker WV, Almubarak H, Long R, Antani S, Frazier SR. Deep Learning Nuclei Detection in Digitized Histology Images by Superpixel. *J Pathol Inform* 2018; 9:10. DOI: 10.4103/jpi.jpi\_74\_17.

#### **IV. AUTOMATED CERVICAL DIGITIZED HISTOLOGY WHOLE SLIDE IMAGE ANALYSIS TOOLBOX**

Sudhir Sornapudi<sup>1</sup>, Ravitej Addanki<sup>1</sup>, R. Joe Stanley<sup>1</sup>, William V. Stoecker<sup>5</sup>, Rodney Long<sup>2</sup>, Rosemary Zuna<sup>3</sup>, Shellaine R. Frazier<sup>4</sup>, Sameer Antani<sup>2</sup>

<sup>1</sup>Department of Electrical and Computer Engineering, Missouri University of Science and Technology, Rolla, MO

<sup>2</sup>Lister Hill National Center for Biomedical Communications for National Library of Medicine, National Institutes of Health, Bethesda, MD

<sup>3</sup>Department of Pathology, University of Oklahoma Health Sciences Center, Oklahoma City, OK

<sup>4</sup>Surgical Pathology Department, University of Missouri Hospitals and Clinics, Columbia, MO

<sup>5</sup>Stoecker & Associates, Rolla MO

#### **ABSTRACT**

Cervical intraepithelial neoplasia (CIN) is regarded as a potential precancerous state of the uterine cervix. Timely and appropriate early treatment of CIN can help reduce cervical cancer mortality. Accurate estimation of CIN grade correlated with human papillomavirus (HPV) type, which is the primary cause of the disease, helps determine the patient's risk for developing the disease. Colposcopy is used to select women for biopsy. Expert pathologists examine the biopsied cervical epithelial tissue under a microscope. The examination can take a long time and is prone to error and often results in high inter- and intra-observer variability in outcomes. We propose a novel image analysis toolbox that can automate CIN diagnosis using whole slide image (digitized biopsies) of cervical tissue samples. The toolbox is built as a four-step deep learning model that detects the epithelium regions, segments the detected epithelial portions,

analyzes local vertical segment regions, and finally classifies each epithelium block with localized attention. We propose an epithelium detection network in this study and make use of our earlier research on epithelium segmentation and CIN classification to complete the design of the end-to-end CIN diagnosis toolbox. The results show that automated epithelium detection and segmentation for CIN classification yields comparable results to manually segmented epithelium CIN classification. This highlights the potential as a tool for automated digitized histology slide image analysis to assist expert pathologists.

**Keywords:** Cervical cancer, cervical intraepithelial neoplasia, detection, segmentation, classification, convolutional neural networks, digital pathology, histology, whole slide image.

## 1. INTRODUCTION

Cervical cancer is widely occurring cancer and a major health problem in women worldwide. It is usually caused by sexually transmitted infections from certain types of Human Papillomavirus (HPV). According to WHO [1], in 2018, cervical cancer was recorded as the second most common cancer in women in low and middle-income regions with an estimated 570,000 new cases and approximately 311,000 deaths occurring during that year [2]. Women aged 20 to 39 years are more vulnerable accounting for 10 premature deaths per week [3]. However, early stage diagnosis can help prevent cervical cancer.

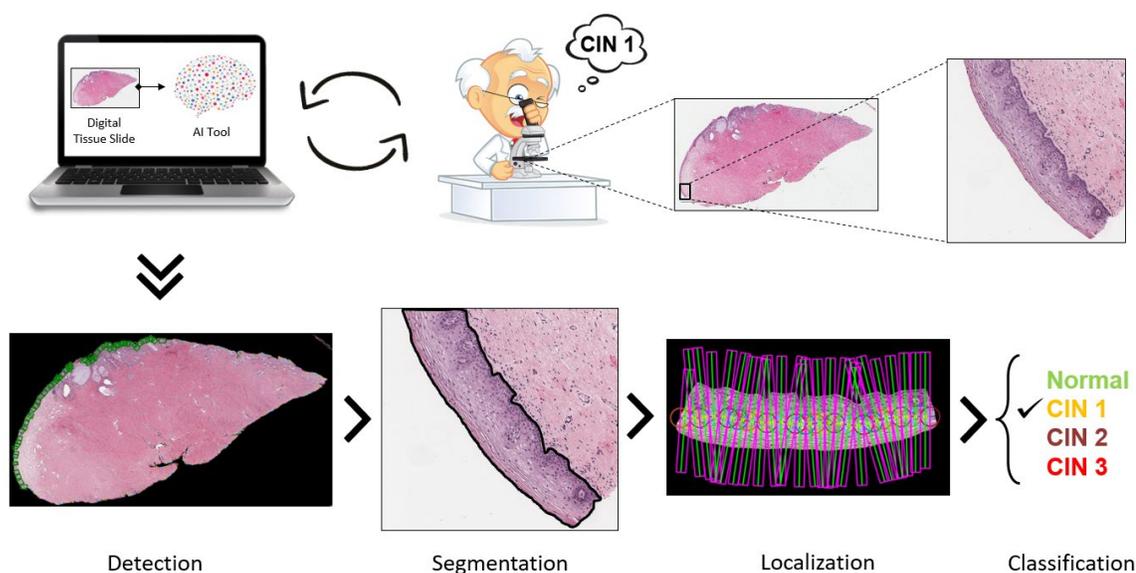


Figure 1. Graphical overview of the proposed toolbox.

Tissue specimens from the uterine cervix of affected women is extracted through biopsy and affixed on glass slides and stained with hematoxylin and eosin (H&E). Then, an expert histopathologist examines the glass sides under a light microscope to provide the diagnosis for each sample, as shown in Figure 1. Accurate interpretation of glass slides is crucial to avoid misdiagnoses [4], which requires extensive time and effort by the pathologist. Each woman could have up to a dozen biopsy samples that require analysis. This displays the necessity of computational digital pathology to augment and automate the process of diagnosis by scanning the digitalized whole slide image (WSI) [5][6].

Grading of cervical disease is largely based on the proportion of immature cells within the epithelium, starting at the base and gradually encompassing the entire epithelial layer. This pre-cancerous condition is called Cervical Intraepithelial Neoplasia (CIN) or cervical dysplasia. The CIN lesions are caused by HPV. All the cells in the

epithelium contain the virus, but the degree to which the epithelium can mature is dependent on the degree to which the virus interferes with the cellular maturation process. The disease is present in the entire epithelial layer in all cases, and the degree of differentiation determines the grade. In CIN1, the immature cells are in the lower one-third; in CIN2, they are found in the lower two-third, and in CIN3, they are found in the entire layer. But the cells at the top are different. They are larger in CIN1 because the tissue can mature to a higher level than the other grades. In CIN3, there is little or no maturation, so the cells at the top look very similar to those at the base [7].

The histopathological WSIs have some unique challenges to overcome. The sheer size of WSI data contains billions of pixels, comprising gigabytes of data. There is a high variability of image appearance due to slide preparation, staining, and various other artefacts during the scanning of the tissue slides. Additionally, the shapes of the biopsied tissue samples vary, and there is no standard shape and size of the epithelium regions and the abnormal cells present inside these regions. The presence of blood stains, ink markers, tapes, and blurred regions pose challenges when designing automated tools. These problems present unique barriers to the development of deep learning models in digital pathology. Nonetheless, the use of deep learning (DL) methods in digital pathology has been proven to have a significant improvement in diagnostic capabilities and efficiency [8][9][10]. The histopathological analysis is performed for various diseases like cervical cancer, skin cancer, breast cancer, prostate cancer, etc. The effects of stain color augmentation and stain color normalization are studied, and an unsupervised approach for stain color normalization was proposed using neural networks for computational pathology [11]. The use of convolution neural networks (CNN) for

segmentation, detection, and classification in common histologic primitives were explored by Janowczyk *et al.* [12]. Multi-instance learning is proposed for image-level classification and annotating relevant regions for histology image analysis [13]. Focusing on cervical cancer, Wang *et al.* [14] presented a block segmentation method to extract textural feature information for CIN classification using support vector machines. Superpixel-based DL nuclei detection was explored in cervical histology images [15]. The problems of inter-observer variability and the advantages of the use of computer-aided systems as a secondary decision for classifying precursor lesions were presented by Albayrak *et al.* [16]. Li *et al.* detailed the use of various machine learning techniques for cervical histopathology image analysis.

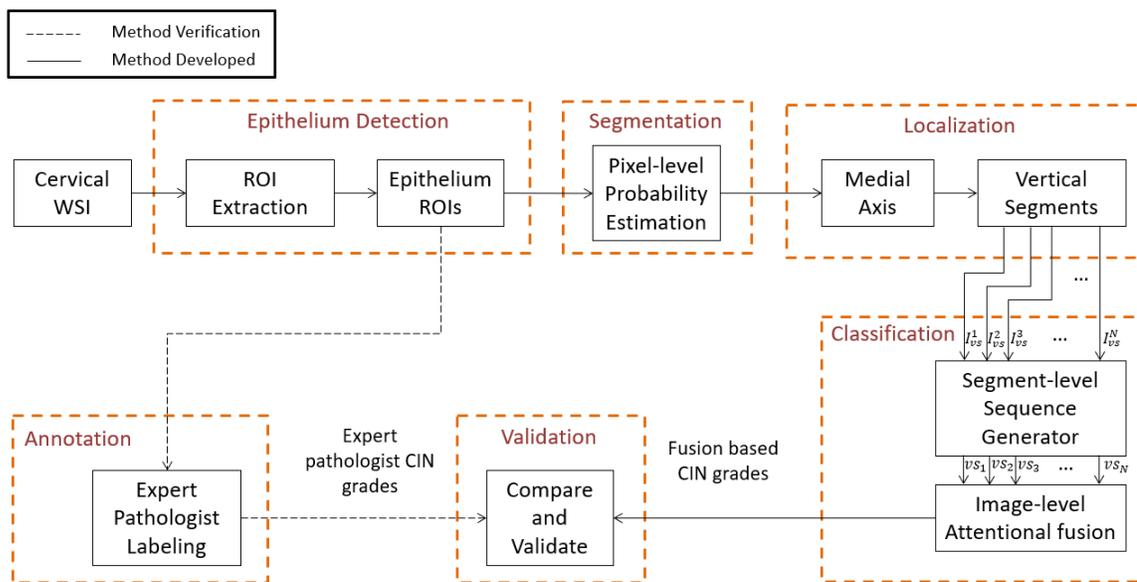


Figure 2. Overview of the proposed toolbox.

Our current study leverages various deep learning models and specifically seeks to automate the diagnosis of cervical cancer by scanning histopathological WSIs. This is

an end-to-end prototype tool that assists pathologists with valuable information like the location of epithelium regions, and can also coarsely segment the epithelium regions from the background and unwanted tissue regions and classify these epithelium regions with added contributions of local regions for the overall classification. We introduce epithelium detection with this study and utilize our previous work on epithelium segmentation [17] and CIN classification [18] to design the toolbox.

This is a novel toolbox that is inspired from the way pathologists analyze the glass slides under a microscope: looking along the outer edges of the tissue and identifying the epithelium regions; zooming in and observing the cell distribution and patterns across the epithelium in detail; and quantifying the CIN grades along the epithelium regions as depicted in Figure 1 and Figure 2.

## **2. METHODOLOGY**

The toolbox incorporates a four-step sequential procedure (Figure 2). First, the outer region of interest (ROI) is identified, and the regions are filtered with the epithelium detection network. Second, pixel-level epithelium segmentation takes place. Third, localization occurs to generate vertical segments. Fourth, CIN grade classification with attention-based sequential feature modeling is completed.

### **2.1. EPITHELIUM DETECTION**

We propose the epithelium detection process with an initial preprocessing that includes the extraction of ROIs from the low-resolution WSIs (refer Section 3.1). This is

followed by a classifier network that identifies the epithelium ROIs by reading the high-resolution versions (refer Section 3.1) of the extracted ROIs.

**2.1.1. ROI Extraction.** Initially, we process the low-resolution version of raw cervical histology WSI to generate a mask for the tissue region, determine the contour, and draw boxes around the outer region of the digitized tissue sample. The WSIs usually have a tissue specimen with a white background. Since the background is uniform, a simple threshold operation can create a mask for the WSI. This mask is further processed to remove small unwanted object regions and close the holes in the object regions.

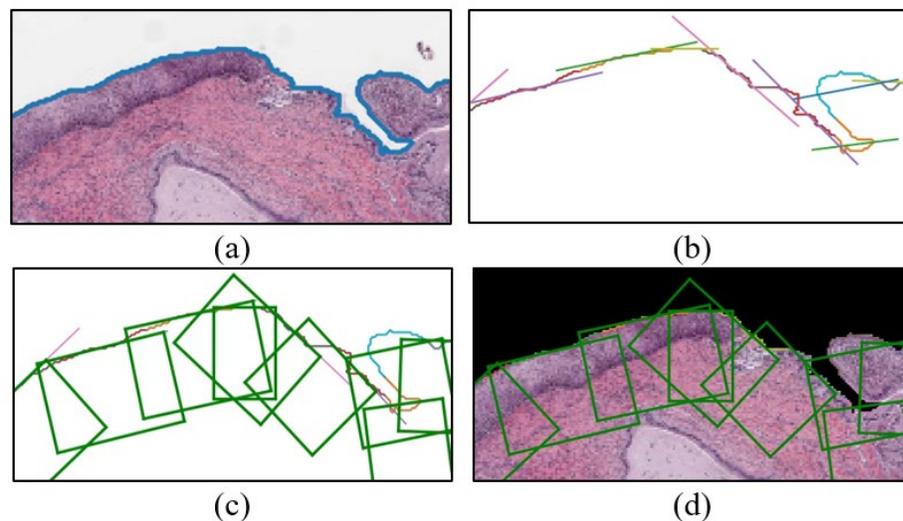


Figure 3. Steps for ROI extraction. (a) Finding the contour on the edge of the tissue sample, (b) Piece-wise curve for drawing tangents, (c) rectangular boxes drawn with reference to tangents, and (d) ROI boxes on the original masked image.

Instead of using grid-based region creation, we optimize the selection of epithelium regions by focusing only on the outer regions, where the epithelium layer is present. The contour of the mask provides the outer edge information. This contour curve

is cut into a piece-wise curve at a frequency of 40 points per cut (chosen empirically based on the low-resolution slide images). In order to draw boxes of ROIs, a polygon is fit based on the points from each piece-wise curve and a tangent is drawn at the midpoint of these piecewise curves. Based on the tangential lines, rectangular boxes were drawn facing the object region of the mask, as shown in Figure 3. The width of the ROI is determined by the maximum and the minimum values of horizontal coordinates and the height is chosen to be 40 pixels (chosen empirically) to accommodate the entire epithelium cross-section. These rectangular box coordinates were normalized and recorded. The high-resolution ROIs (at 10 $\times$  magnification) were finally created by cropping out the image regions from the high-resolution slide image using the normalized rectangular bounding box coordinates data as shown in Figure 4.

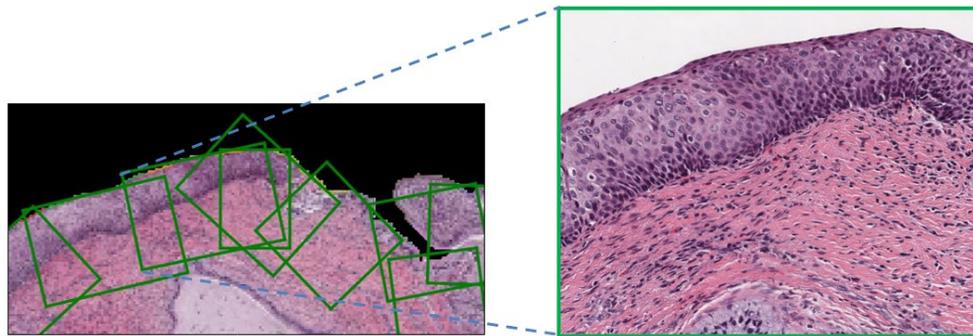


Figure 4. Mapping of high-resolution ROI (right) to its low-resolution image (left).

**2.1.2. Epithelium Detection Network.** The epithelium detection network is a binary classifier that categories an input image as epithelium or non-epithelium. The high-resolution ROIs are fed to this network to filter and retain only the epithelium containing ROIs, as shown in Figure 5.

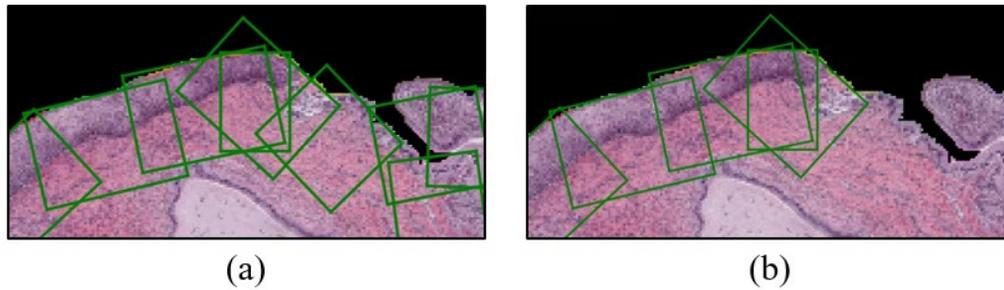


Figure 5. Filtering of epithelium ROIs with the results from the epithelium detection network.

Table 1 presents the network architecture that was investigated. This is a customized version of the VGG-19 model [19], where the initial layers contain a series of convolution block and max-pooling layers. The end feature maps generated from these layers are vectorized and passed through fully connected layers. All the aforementioned layers were activated with ReLU non-linearity functions, except the last fully connected layer that contains two neurons to compute the classification probability for each class using the SoftMax activation function. To reduce overfitting, the output of the first fully connected layer is constrained by randomly dropping 50% of the values to zero.

The weights in the convolutional layers are initialized with Kaiming initialization [20] for better stability, and the fully connected layers are initialized with the normal distribution. In the training phase, the weights are iteratively updated with the gradients of the cross-entropy loss function, which is computed via RMSprop optimization over a mini-batch of training samples. The initial learning rate is set to 0.0001 and changes adaptively as the training progresses.

Table 1. Epithelium detection network architecture.

Layers	Configurations	Size
Input	-	$3 \times 250 \times 250$
Conv Block 1	$[k: 3 \times 3, s: 1, p: 1] \times 2$	$64 \times 250 \times 250$
Pool 1	$mp: 2 \times 2, s: 2$	$64 \times 125 \times 125$
Conv Block 2	$[k: 3 \times 3, s: 1, p: 1] \times 2$	$128 \times 125 \times 125$
Pool 2	$mp: 2 \times 2, s: 2$	$128 \times 62 \times 62$
Conv Block 3	$[k: 3 \times 3, s: 1, p: 1] \times 4$	$256 \times 62 \times 62$
Pool 3	$mp: 2 \times 2, s: 2$	$256 \times 31 \times 31$
Conv Block 4	$[k: 3 \times 3, s: 1, p: 1] \times 4$	$512 \times 31 \times 31$
Pool 4	$mp: 2 \times 2, s: 2$	$512 \times 15 \times 15$
Conv Block 5	$[k: 3 \times 3, s: 1, p: 1] \times 4$	$512 \times 15 \times 15$
Pool 5	$mp: 2 \times 2, s: 2$	$512 \times 7 \times 7$
Flatten	-	$25088 \times 1$
FC 1	$nh: 1024$	$1024 \times 1$
Dropout	$prob: 0.5$	$1024 \times 1$
FC 2	$nh: 1024$	$1024 \times 1$
FC 3	$nh: 2$	$2 \times 1$
Output	$softmax$	$2 \times 1$

$k$ ,  $s$ ,  $p$ ,  $mp$ ,  $nh$ ,  $prob$  are kernel, stride size, padding size, max pooling, number of neurons, and probability, respectively. ‘FC’ denotes a fully connected single-layer neural network.

## 2.2. EPITHELIUM SEGMENTATION

From Figure 2, epithelium segmentation is the second step in the slide analysis process. We utilize the EpithNet model from our previous studies [17] to coarsely segment the high-resolution epithelium ROIs to generate an epithelium segmentation mask. The segmentation model is a pixel-wise epithelial probability estimator and is developed based on the information provided by a pixel depending on the surrounding spatial proximity in the image plane. The epithelium ROI is preprocessed by splitting into tiles, and each tile is further processed to generate  $64 \times 64 \times 3$  RGB patch image data. These patches are created with a sliding window technique with stride 4. From [17], the EpithNet-64 regression model was utilized to process these patch data to output an

estimated probability of the center pixel of being an epithelium. These pixel probabilities are gathered and treated as pixel intensities to finally form a mask. This mask is post-processed by applying thresholding, morphology, and smoothing filters to finally generate a binary segmentation mask.

### 2.3. LOCALIZATION

CIN is the growth of atypical cervical cells in the epithelium. This abnormal growth is clearly understood when observed locally. Thus, standard width vertical segments [18] are generated from the epithelium ROIs with reference to the medial axis, drawn with the help of epithelium segmentation mask information, as shown in Figure 6. The details are provided in our previous work [18].

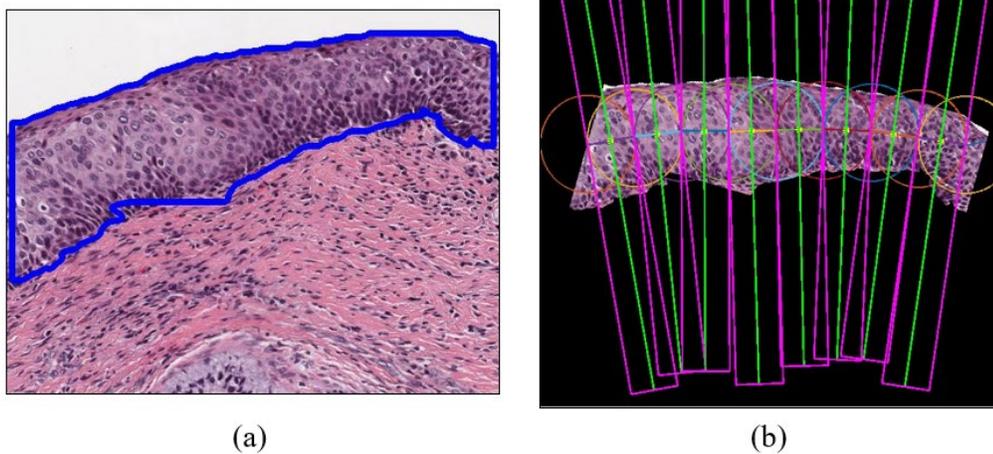


Figure 6. (a) Epithelium segmentation mask overlaid as a contour on the epithelium ROI.  
(b) Vertical segments generation through the localization process.

## 2.4. CIN CLASSIFICATION

For each segmented epithelium, CIN classification is performed. The DeepCIN [18] is a two-fold learning process. First, a segment-level sequence generator is a weakly supervised network that scans each localized vertical segment image to generate the best sequence representation of the input image. This is built as an encoder-decoder model, where the encoder is a CNN model that extracts and encodes convolutional spatial feature information to a sequential feature. The decoder is a many-to-one model that consists of two layers of bidirectional Long-Short-Term Memory (BLSTM) network and a single layer neural network. Second, an attention-based fusion network is an image-level classifier that sequentially interprets the vertical segment sequences. This provides a contextual understanding of local information that not only helps in providing the multi-class CIN classification result, but also provides the contribution of each vertical segment towards the final classification. This is built with gated recurrent units (GRUs) and attentional neural network layers. The detailed model implementation can be found in the work of Sornapudi *et al.* [18].

The model was previously trained with 5-fold cross-validation, and we ensemble the five trained models to produce the CIN classification result on the proposed toolbox.

## 3. EXPERIMENTS

### 3.1. DATA

The study uses 150 H&E stained cervical histopathological slides (WSI) provided by the Department of Pathology at the University of Oklahoma Medical Center in

collaboration with the National Library of Medicine. The glass slides were scanned using the Aperio ScanScope slide scanner with 20× objective, producing WSIs in a pyramidal tiled format with the file extension “svs”. These SVS files are large (they typically range from 1 gigabyte to 100 megabytes). Each pixel has a size of 0.25  $\mu\text{m}^2$ . The pyramidal tile level varies from 0 to 2/3/4. In this paper, we often refer to a 1× magnification image (highest pyramid level) as a low-resolution image and 20× magnification image (pyramid level 0) down sampled to 10× magnification as a high-resolution image. This is explicitly performed to maintain the same image resolutions used in our previous works [17][18]. There are three sets of WSIs captured during the years 2013, 2015, and 2016, and hence named OU13, OU15, and OU16, respectively. Each of these sets contains 50 WSIs. The study uses 50 WSIs from the OU13 set for training and validation of the epithelium detection model, and 100 WSIs from both the OU15 set and the OU16 set for testing our toolbox. With our automated ROI extraction technique, we could generate high-resolution arbitrary size images that contain epithelium and non-epithelium regions. The distribution of the image blocks can be observed in Table 2. The images from Table 2 will be evaluated for the correctness of the epithelium segmentation process in Section 4.1.

Table 2. Data distribution for epithelium detection.

Dataset	WSIs	Epithelium ROIs	Non-epithelium ROIs
OU13	50	2,998	20,841
OU15	50	4,915	12,595
OU16	50	4,106	8,601

The data set examined in this research for evaluating CIN classification model (Figure 2) consists of a total of 947 expert-labeled epithelium images (a subset of obtained epithelium ROIs), 723 images from OU15-set, and 224 images from OU16-set. The class distribution of the data is shown in Table 3. It should be noted that these 947 epithelium images are an independent set of images extracted from the proposed approach and are mutually exclusive from the manually extracted epithelium images that are used for training the CIN classification model [18].

Table 3. Subset of epithelium ROI images for evaluating CIN classification.

Class	OU15	OU16	Combined set
Normal	451	133	584
CIN 1	90	11	101
CIN 2	128	41	169
CIN 3	54	39	93
Total	723	224	947

### 3.2. IMPLEMENTATION DETAILS

The architecture of the epithelium detection network is summarized in Table 1. The network incorporates a transfer learning scheme. So, the entire model is pre-trained on the ImageNet classification dataset. The convolutional module weights were frozen, and the rest of the layer weights were reinitialized with random Gaussian distributions.

We have designed the CNN model such that it can read RGB input images of size  $250 \times 250 \times 3$ . To maintain a standard resolution of the input images, the extracted ROIs are padded with zeros, center cropped to size  $500 \times 500 \times 3$  and finally resized to  $250 \times 250 \times 3$ . We incorporated data augmentation techniques to avoid the problem of highly

imbalanced data in the training set (Table 2). The epithelium ROI images are upsampled to count equivalent to non-epithelium ROIs (20,841) via augmentations like random rotate, vertical and horizontal flip, random blur, etc. RMSprop, with a mini-batch of size 32, is used to train the network for 100 epochs. Early stopping is applied to monitor the generalization error and avoid overfitting.

In the testing phase, the ROIs categorized as epithelium were further processed with previously trained models in the toolbox: EpithNet-64 and DeepCIN. We obtain an epithelium segmentation mask with EpithNet-64 for the generation of vertical segments that are consumed by DeepCIN to deliver the CIN classification results and the contribution of the vertical segments towards the classification output. The models are run on the PyTorch v1.4 platform [21] using Nvidia Quadro P4000 GPU with 8GB of memory.

### 3.3. EVALUATION METRICS

We evaluate the proposed epithelium detection network for classification as epithelium/non-epithelium ROIs. The performance evaluation metrics include specificity ( $Sp$ ), sensitivity ( $Se$ ), harmonic mean ( $H_{mean}$ ), F1-score ( $F1$ ), accuracy ( $ACC$ ), and area under the ROC curve ( $AUC$ ).  $Sp$  measures the proportion of correctly identified non-epithelium ROIs,  $Se$  measures the proportion of correctly identified epithelium ROIs,  $H_{mean}$  is the harmonic mean of  $Sp$  and  $Se$  (which is better at measuring under imbalance data distribution),  $F1$  is the harmonic mean of precision and recall, and  $ACC$  is the global accuracy.  $AUC$  is the area under the receiver operating characteristic curve and is plotted with varying thresholds on final classification scores.

We also evaluate the final CIN classification results from the detected epithelium ROIs. The scoring metrics used are precision ( $P$ ), recall ( $R$ ), F1-score ( $F1$ ), classification accuracy ( $ACC$ ), area under Receiver Operating Characteristic curve ( $AUC$ ), average precision ( $AP$ ), Matthews correlation coefficient ( $MCC$ ), and Cohen’s kappa score ( $\kappa$ ) [18]. The percentage weighted average scores were computed to account for the imbalance in the data distribution.

## 4. RESULTS

We evaluate the toolbox performance by comparing the epithelium detection network results and the CIN classification results against the expert pathologist annotated ground truths on the OU15 and OU16 WSI datasets.

### 4.1. PERFORMANCE OF EPITHELIUM DETECTION NETWORK

Table 4 shows the classification performance ( $Sp$ ,  $Se$ ,  $H_{mean}$ ,  $F1$ ,  $ACC$ , and  $AUC$ ) of the proposed epithelium detection network. The objective of this network is to sort the extracted ROI images into epithelium and non-epithelium. Since there are more non-epithelium ROIs compared to epithelium ROIs (Table 2), the specificity is always observed to be higher than sensitivity. Harmonic mean ( $H_{mean}$ ) gives a better-balanced score between  $Sp$  and  $Se$ , and is found to have a mean value of 97.3%, 92.7%, and 95.0% among OU15, OU16, and OU15 and OU16 combined datasets, respectively. We observed that the trained network has better generalization on the OU15-set, compared to the OU16-set. The combined dataset results were also reported. We could not compare

the performance of the network with other works because, to our knowledge, this is the first study on cervical epithelium detection.

Table 4. Epithelium detection results.

Test set	$Sp$	$Se$	$H_{mean}$	$F1$	$ACC$	$AUC$
OU15	98.3	96.6	97.3	95.6	97.8	97.4
OU16	96.3	90.8	92.7	91.4	95.2	93.5
OU15/OU16	97.3	93.7	95.0	93.5	96.5	95.5

Figure 7 contains examples of correctly classified epithelium ROIs (true positive) and misclassified epithelium ROIs (false positive). Typically, the cancer cells are manifested in the epithelium, and hence the identification of epithelium is our top priority. The network is observed to identify the epithelium regions even under challenging conditions. The falsely identified ROIs closely resemble the epithelium regions, which makes the classification task difficult. Nevertheless, the network has provided good performance accuracy results of 97.8% on OU15-set, 95.2% on OU-16, and 96.5% on the combined set.

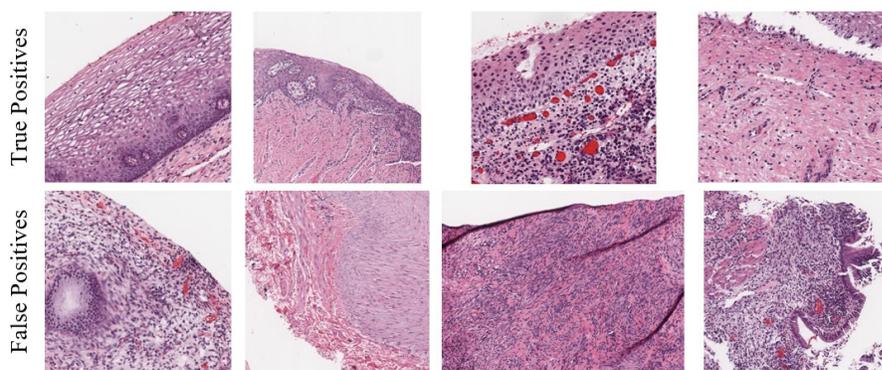


Figure 7. Examples of epithelium detection results. Correctly classified (top row) and misclassified (bottom row) epithelium ROIs.

## 4.2. PERFORMANCE OF CIN CLASSIFICATION MODEL

We evaluate and compare the performance of the CIN classification model on the high-resolution epithelium images extracted through the proposed automated epithelium detection and segmentation process, and manually cropped and segmented images. We employ five scoring schemes [18] to analyze the classification results. They are exact class labels, CIN versus Normal, CIN3-CIN2 versus CIN1-Normal, CIN3 versus CIN2-CIN1-Normal, and off-by-one class.

Table 5. CIN classification results on OU15-set.

Scoring Scheme	P	R	F1	ACC	AUC	AP	MCC	$\kappa$
Exact class label	83.1	83.8	82.8	83.8	94.4	86.8	70.35	70.1
CIN vs Normal	91.1	91.1	91.1	91.1	90.1	95.7	81.0	81.0
CIN3-CIN2 vs CIN1-Normal	93.2	93.2	93.8	93.2	89.1	97.8	81.6	81.3
CIN3 vs CIN2-CIN1-Normal	93.6	94.2	92.8	94.2	63.7	95.5	46.1	39.4
Off-by-one	-	-	-	96.3	-	-	-	-

Table 6. CIN classification results on OU16-set.

Scoring Scheme	P	R	F1	ACC	AUC	AP	MCC	$\kappa$
Exact class label	90.2	88.4	88.2	88.4	98.0	93.1	80.5	80.0
CIN vs Normal	97.3	97.3	97.3	97.3	97.2	99.7	94.4	94.4
CIN3-CIN2 vs CIN1-Normal	95.7	95.6	95.5	95.5	94.0	99.1	90.3	90.0
CIN3 vs CIN2-CIN1-Normal	93.0	92.4	91.5	92.4	78.2	97.0	71.9	68.1
Off-by-one	-	-	-	98.2	-	-	-	-

The classification results from the DeepCIN classification model for OU15 and OU16 image sets are tabulated in Table 5 and Table 6, respectively. The results indicate that the DeepCIN model performed better on the OU16 dataset compared to the OU15 dataset. This may be due to the presence of relatively fewer artefacts during the preparation of the OU16 WSIs compared to OU15 WSIs. Table 7 shows the results of the combined dataset. We observe that the model has few misclassifications, usually off by one CIN grade. This can be observed by the off-by-one class results. This disagreement is also observed to happen among expert pathologists with interobserver variability. If we rank the scoring schemes based on the results, the off-by-one class is followed by CIN versus Normal (abnormal vs normal), which is helpful to distinguish the abnormal precancerous epithelium regions from the normal epithelium regions. These are followed by, CIN3-CIN2 versus CIN1-Normal, CIN3 versus CIN2-CIN1-Normal, and exact class labels.

The proposed toolbox is observed to face difficulty in correctly identifying the CIN 3 epithelium images. There is an off-by-one grade error with misclassification as CIN 2. This can be observed from the metric values of CIN3-CIN2 versus CIN1-Normal and CIN3 versus CIN2-CIN1-Normal scoring schemes in Table 7.

The performance of the proposed toolbox for automated cervical diagnosis is benchmarked against CIN classification results on the manually cropped and segmented epithelium images (Table 8). The manually extracted epithelium images were chosen carefully to capture and focus on the epithelium regions along with accurate annotations for epithelium masks. These images are close to the ideal conditions, and we compare them with the epithelium images from an automated realistic toolbox. We observed that

the proposed toolbox has a closer performance to the benchmark results, and this indicates that the proposed prototype has the potential to be used in real-world clinical settings.

Table 7. CIN classification results on the combined set.

Scoring Scheme	P	R	F1	ACC	AUC	AP	MCC	K
Exact class label	85.0	85.0	84.2	85.0	95.5	88.3	73.0	72.7
CIN vs Normal	92.6	92.6	92.6	92.6	92.0	96.9	84.3	84.3
CIN3-CIN2 vs CIN1-Normal	93.8	93.8	93.7	93.8	90.5	98.3	84.1	83.9
CIN3 vs CIN2-CIN1-Normal	93.7	93.8	92.6	93.8	69.7	96.0	58.3	52.9
Off-by-one	-	-	-	96.7	-	-	-	-

Table 8. Benchmark CIN classification results [18].

Scoring Scheme	P	R	F1	ACC	AUC	AP	MCC	K
Exact class label	88.6	88.5	88.0	88.5	96.5	91.5	82.0	81.5
CIN vs Normal	94.6	94.1	94.0	94.1	93.8	97.7	88.5	87.9
CIN3-CIN2 vs CIN1-Normal	96.8	96.7	96.7	96.7	96.0	98.9	92.7	92.5
CIN3 vs CIN2-CIN1-Normal	96.2	96.0	96.0	96.0	88.4	98.3	85.3	84.8
Off-by-one	-	-	-	98.9	-	-	-	-

## 5. DISCUSSION

The cervical histopathology data suffers from three major limitations in the data collection and preparation. Firstly, unlike scenic images from various public challenges,

biomedical image data requires a lot of approvals to gather patient data and hence the amount of data is relatively very low. Secondly, gathering expert labelled data is always challenging since this requires skills to identify the regions and grade the cancer. There is always an inter-pathologist variation related to interpretation of the results. A study [22] has shown that there is an interobserver variability of 0.799 to 0.887 in terms of kappa score among four expert pathologists who have CIN grading experience of 8-30 years. Thirdly, the distribution of the data is always skewed/ imbalanced as shown in Table 2 and Table 3.

This paper is intended to compare the proposed automated digitized histology slide analysis for CIN classification of epithelium regions with the manually segmented epithelium regions. Fewer epithelium regions in WSIs were considered for evaluation due to limited availability of expert pathologist labeling. Future studies will explore the interpretation of WSI-level CIN classification for the complete end-to-end digitized slide analysis. The inclusion of techniques like graph theory for deeper understanding of spatial context and data fusion might help in further improving the CIN classification results. Stacked models can be created to handle the lack of consensus pathology, that is, the designed models should have the ability to interpret the disagreements among the pathologists' ground truth labelling. The resolution of WSI scanners should be a concern too. There is variability across manufacturers which leads to issues with different image resolutions. The future work will be also be focused on designing models that can handle WSIs from various sources.

## 6. CONCLUSION

Our pipeline draws inspiration from the examination strategy of an expert pathologist, where he/she scrutinizes the growth of abnormal cells across small portions of the epithelium. This is realized by scanning the cervical histopathological WSI and extracting the epithelium ROIs present on the outer layer of the tissue sample. Since there are regions without epithelium, filtering the ROIs is crucial to retain only the epithelium ROIs and this is accomplished by the proposed epithelium detection network. With the help of our previous studies, we incorporated the EpithNet-64 model for segmenting the epithelium regions in the epithelium ROIs. Small vertical portions are extracted for a localized cell growth pattern analysis, which is performed by the DeepCIN model. The results sequences are fused with attentional observation to determine the final CIN grade for the epithelium ROI. Even the significance of the local regions was identified in this process of CIN classification. Furthermore, the CIN grade for the entire WSI can be generated by voting CIN classification results from the portions of epithelium ROIs.

We observed that our unique, novel approach for an automated CIN diagnosis from a WSI has achieved expert pathologist level accuracy. This clearly indicates the potential of our proposed pipeline as an assisting tool to an expert pathologist both in terms of quality of diagnosis and time. Due to the limitation of the data samples and expert annotated WSI-level labels, we tend to quote the toolbox as a prototype. If there is the availability of more data from various sources, the toolbox could be better generalized for use by everyone. The tool can be further improved by considering additional information of patients' metadata and genetic codes.

## ACKNOWLEDGMENTS

This research was supported in part by the Intramural Research Program of the National Institutes of Health (NIH) [grant number 10.13039/100000002]; National Library of Medicine (NLM) [grant number 10.13039/100000092]; and Lister Hill National Center for Biomedical Communications (LHNCBC). In addition, we gratefully acknowledge the medical expertise of Dr. P. Deepika Devada.

## REFERENCES

- [1] Human papillomavirus (HPV) and cervical cancer. World Health Organization; 2019. Available from: [https://www.who.int/news-room/fact-sheets/detail/human-papillomavirus-\(hpv\)-and-cervical-cancer](https://www.who.int/news-room/fact-sheets/detail/human-papillomavirus-(hpv)-and-cervical-cancer). [Last accessed: 2020 Apr 29].
- [2] Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I. Global Cancer Observatory: Cancer Today. Lyon, France, 2018.
- [3] Siegel RL, Miller KD, and Jemal A. Cancer statistics, 2020. *CA Cancer J Clin* 2020; 70: 7–30.
- [4] Olhoffer IH, Lazova R, and Leffell DJ. Histopathologic Misdiagnoses and Their Clinical Consequences. *Arch Dermatol* 2002; 138(10): 1381–1383.
- [5] Gage JC, Joste N, Ronnett BM, Stoler M, Hunt WC, Schiffman M *et al.* A comparison of cervical histopathology variability using whole slide digitized images versus glass slides: experience with a statewide registry. *Hum Pathol* 2013; 44(11):2542–2548.
- [6] Bongaerts O, Clevers C, Debets M, Paffen D, Senden L, Rijks K *et al.* Conventional Microscopical versus Digital Whole-Slide Imaging-Based Diagnosis of Thin-Layer Cervical Specimens: A Validation Study. *J Pathol Inform* 2018; 9:29.
- [7] Melnikow J, Nuovo J, Willan AR, Chan BKS and Howell LP. Natural History of Cervical Squamous Intraepithelial Lesions: A Meta-Analysis. *Obstet Gynecol* 1998; 92.

- [8] Litjens G, Sánchez CI, Timofeeva N, Hermsen M, Nagtegaal I, Kovacs I *et al.* Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Sci Rep* 2016; 6.
- [9] Ianni JD, Soans RE, Sankarapandian S, Chamarthi RV, Ayyagari D, Olsen TG *et al.* Tailored for Real-World: A Whole Slide Image Classification System Validated on Uncurated Multi-Site Data Emulating the Prospective Pathology Workload. *Sci Rep* 2020; 10(1):3217.
- [10] Sornapudi S. Nuclei segmentation of histology images based on deep learning and color quantization and analysis of real world pill images. 2017; p. 40.
- [11] Tellez D, Litjens G, Bándi P, Bulten W, Bokhorst JM, Ciompi F, *et al.* Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. *Med Image Anal* 2019; 58.
- [12] Janowczyk A and Madabhushi A. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *J Pathol Inform* 2016; 7:29.
- [13] Vu T, Lai P, Raich R, Pham A, Fern XZ, and Rao UA. A Novel Attribute-based Symmetric Multiple Instance Learning for Histopathological Image Analysis. *IEEE Trans Med Imaging* 2020.
- [14] Wang Y, Crookes D, Eldin OS, Wang S, Hamilton P, and Diamond J. Assisted diagnosis of cervical intraepithelial neoplasia (CIN). *IEEE J Sel Top Signal Process* 2009; 3: 112–121.
- [15] Sornapudi S, Stanley RJ, Stoecker WV, Almubarak H, Long R, Antani S, *et al.* Deep Learning Nuclei Detection in Digitized Histology Images by Superpixels. *J Pathol Inform* 2018; 9:5.
- [16] Albayrak A, Unlu A, Calik N, Capar A, Bilgin G, Toreyin BU *et al.* A Whole Slide Image Grading Benchmark and Tissue Classification for Cervical Cancer Precursor Lesions with Inter-Observer Variability. *ArXiv* 2018; pp. 1–15.
- [17] Sornapudi S, Hagerty J, Stanley R J, Stoecker WV, Long R, Antani S, *et al.* EpithNet: Deep regression for epithelium segmentation in cervical histology images. *J Pathol Inform* 2020; 11:10.
- [18] Sornapudi S, Stanley RJ, Stoecker WV, Long R, Xue Z, Zuna R. *et al.* Feature based Sequential Classifier with Attention Mechanism. *ArXiv: 2007.11392 [eess.IV]* 2020.
- [19] Simonyan K and Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. 2014.

- [20] He K, Zhang X, Ren S, and Sun J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. Proc IEEE Int Conf Comput Vis 2015; p. 1026–1034.
- [21] Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, et al. Automatic differentiation in PyTorch. 2017.
- [22] Cai B, Ronnett B M, Stoler M, Ferenczy A, Kurman R J, Sadow D, et al. Longitudinal Evaluation of Interobserver and Intraobserver Agreement of Cervical Intraepithelial Neoplasia Diagnosis Among an Experienced Panel of Gynecologic Pathologists. Am J Surg Pathol 2007; 31(12).

## SECTION

### 2. SUMMARY AND CONCLUSIONS

This dissertation proposes deep learning models for uterine cervical histology image analysis. The analysis was performed by applying CNN models for nuclei detection, epithelium segmentation, CIN classification, and epithelium detection. The latter three works were successfully combined to design a prototype tool that can automatically diagnose the CIN from the cervical histology WSIs. There are tools that can automate the abnormality and cancer detection from various histology datasets, but there is no tool that can specifically address cervical cancer. This work fills that void, and from the results, it can be observed that the proposed toolbox has the potential to address the cervical cancer problem in health care. This prototype tool can be scaled up by retraining it with cervical histology image datasets from various sources so that the tool can be generically used for real-world clinical purposes around the world. This can be helpful in multiple ways such as use as an assistance tool for an expert pathologist, use as a second opinion, or even as a virtual pathologist where there is a scarcity of pathologists to review the histology images.

To conclude, deep learning has proven to produce results that can clearly surpass the state-of-the-art methods in comprehensive image analysis, decision-making, and enhanced classification of cervical histopathology. The performance accuracy of the models is comparable to the expert pathologist-level accuracies.

## BIBLIOGRAPHY

- [1] J. Niederhuber, J. Armitage, J. Doroshow, M. Kastan, and J. Tepper, “Cancers of the cervix, vulva and vagina,” in *Abeloff’s Clinical Oncology*, 6th ed., Philadelphia, PA: Elsevier, 2020.
- [2] “Key Statistics for Cervical Cancer,” *American Cancer Society*, 2020. [Online]. Available: <https://www.cancer.org/cancer/cervical-cancer/about/key-statistics.html>. [Accessed: 07-May-2020].
- [3] B. F. Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I, “Global Cancer Observatory: Cancer Today,” Lyon, France, 2018.
- [4] World Health Organization, “Sexual and reproductive health, Prevention and Control of Cervical cancer,” 2019. [Online]. Available: <https://www.who.int/reproductivehealth/topics/cancers/en/>. [Accessed: 07-May-2020].
- [5] AskMayoExpert., “Cervical cancer (adult),” *Mayo Foundation for Medical Education and Research*, 2019. .
- [6] PDQ Adult Treatment Editorial Board, *Cervical Cancer Treatment (PDQ®): Health Professional Version*. Bethesda, MD: National Cancer Institute (US), 2020.
- [7] K. Maniar and J. Wei, “Pathology of Cervical Carcinoma,” in *The Global Library of Women’s Medicine*, 2017.
- [8] W. G. McCluggage *et al.*, “Inter- and intra-observer variation in the histopathological reporting of cervical squamous in traepithelial lesion using a modified Bethesda grading system,” *BJOG An Int. J. Obstet. Gynaecol.*, vol. 105, no. 2, pp. 206–210, 1998.
- [9] A. Madabhushi and G. Lee, “Image analysis and machine learning in digital pathology: Challenges and opportunities,” *Med. Image Anal.*, vol. 33, pp. 170–175, 2016.
- [10] C. Li *et al.*, “Cervical Histopathology Image Classification Using Multilayer Hidden Conditional Random Fields and Weakly Supervised Learning,” *IEEE Access*, vol. 7, pp. 90378–90397, 2019.
- [11] A. J. Schaumberg *et al.*, “Large-Scale Annotation of Histopathology Images from Social Media,” *bioRxiv*, 2018.

- [12] L. Wei, Q. Gan, and T. Ji, "Cervical cancer histology image identification method based on texture and lesion area features.," *Comput. Assist. Surg. (Abingdon, England)*, vol. 22, no. sup1, pp. 186–199, Dec. 2017.
- [13] C. Li *et al.*, "Cervical Histopathology Image Clustering Using Graph Based Unsupervised Learning," in *Proceedings of the 11th International Conference on Modelling, Identification and Control (ICMIC2019)*, 2020, pp. 141–152.
- [14] S. Sornapudi *et al.*, "Deep Learning Nuclei Detection in Digitized Histology Images by Superpixels," *J. Pathol. Inform.*, vol. 9, no. 1, p. 10, 2018.
- [15] C. Li *et al.*, "Transfer Learning Based Classification of Cervical Cancer Immunohistochemistry Images," in *Proceedings of the Third International Symposium on Image Computing and Digital Medicine*, 2019, pp. 102–106.
- [16] C. Li *et al.*, "Cervical Histopathology Image Classification Using Ensembled Transfer Learning," in *International conference on information technologies in biomedicine*, 2019, pp. 26–37.
- [17] F. Sheikhzadeh, "Improving cervical neoplasia diagnosis via novel in vivo imaging technologies and deep learning algorithms," University of British Columbia, 2018.
- [18] H. A. Almubarak *et al.*, "Convolutional Neural Network Based Localized Classification of Uterine Cervical Cancer Digital Histology Images.," *Procedia Comput. Sci.*, vol. 114, pp. 281–287, 2017.
- [19] S. J. Keenan *et al.*, "An automated machine vision system for the histological grading of cervical intraepithelial neoplasia (CIN).," *J. Pathol.*, vol. 192, no. 3, pp. 351–62, 2000.
- [20] B. Weyn, W. A. A. Tjalma, P. Vermeylen, A. van Daele, E. Van Marck, and W. Jacob, "Determination of tumour prognosis based on angiogenesis-related vascular patterns measured by fractal and syntactic structure analysis.," *Clin. Oncol. (R. Coll. Radiol.)*, vol. 16, no. 4, pp. 307–316, Jun. 2004.
- [21] F. Sheikhzadeh *et al.*, "Confocal fluorescence microscopy for detection of cervical preneoplastic lesions," in *Proc.SPIE*, 2015, vol. 9420.
- [22] Y. Wang, D. Crookes, O. S. Eldin, S. Wang, P. Hamilton, and J. Diamond, "Assisted diagnosis of cervical intraepithelial neoplasia (CIN)," *IEEE J. Sel. Top. Signal Process.*, vol. 3, no. 1, pp. 112–121, 2009.
- [23] G. H. B. Miranda, J. Barrera, E. G. Soares, and J. C. Felipe, "Structural Analysis of Histological Images to Aid Diagnosis of Cervical Cancer," in *2012 25th SIBGRAPI Conference on Graphics, Patterns and Images*, 2012, pp. 316–323.

- [24] F. Sheikhzadeh *et al.*, “Quantification of confocal fluorescence microscopy for the detection of cervical intraepithelial neoplasia.,” *Biomed. Eng. Online*, vol. 14, p. 96, Oct. 2015.
- [25] J. Saltz *et al.*, “Spatial Organization and Molecular Correlation of Tumor-Infiltrating Lymphocytes Using Deep Learning on Pathology Images.,” *Cell Rep.*, vol. 23, no. 1, pp. 181-193.e7, Apr. 2018.
- [26] S. Sornapudi *et al.*, “EpithNet: Deep regression for epithelium segmentation in cervical histology images,” *J. Pathol. Inform.*, vol. 11, no. 1, p. 10, 2020.
- [27] P. Guo *et al.*, “Nuclei-Based Features for Uterine Cervical Cancer Histology Image Analysis With Fusion-Based Classification,” *IEEE J. Biomed. Heal. Informatics*, vol. 20, no. 6, pp. 1595–1607, 2016.
- [28] P. Guo *et al.*, “Enhancements in localized classification for uterine cervical cancer digital histology image assessment.,” *J. Pathol. Inform.*, vol. 7, p. 51, 2016.
- [29] S. De *et al.*, “A fusion-based approach for uterine cervical cancer histology image classification.,” *Comput. Med. Imaging Graph.*, vol. 37, no. 7–8, pp. 475–87, 2013.
- [30] M. Jondet, R. Agoli-Agbo, and L. Dehennin, “Automatic measurement of epithelium differentiation and classification of cervical intraneoplasia by computerized image analysis,” *Diagn. Pathol.*, vol. 5, p. 7, Jan. 2010.
- [31] S. Sornapudi *et al.*, “Feature based Sequential Classifier with Attention Mechanism,” ArXiv: 2007.11392 [eess.IV] 2020.
- [32] C. Li *et al.*, “A review for cervical histopathology image analysis using machine vision approaches,” *Artif. Intell. Rev.*, 2020.
- [33] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, “SLIC superpixels compared to state-of-the-art superpixel methods,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2281, 2012.

## VITA

Sudhir Sornapudi was born in Visakhapatnam, Andhra Pradesh, India. Before attending Missouri University of Science and Technology, Rolla, USA, he attended JNTUK – University College of Engineering Vizianagaram, India, where he earned a Bachelor of Technology degree in Electronics and Communications Engineering in 2014. He then pursued his Master of Science (thesis) in Computer Engineering at Missouri University of Science and Technology in May 2017. He received his Ph.D. degree in Computer Engineering from Missouri University of Science and Technology in December 2020. His academic experience included working as a Graduate Research Assistant and Graduate Teaching Assistant. He also had industrial experience interning at various companies like MilliporeSigma, Merck KGaA; National Library of Medicine (NLM), NIH; Xyken LLC; and Naval Science and Technological Laboratory (NSTL), DRDO. His interests mainly included computer vision, image processing/analysis, machine learning, deep learning, and embedded systems.