

RICE UNIVERSITY

By

Brady Hunt

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE

Rebecca Richards-Kortum

Rebecca Richards-Kortum

Kathleen M. Schmeler
Kathleen M. Schmeler (May 18, 2020)

Kathleen Schmeler

Tomasz Tkaczyk

Tomasz Tkaczyk (May 18, 2020)

Tomasz Tkaczyk

Philip Castle

Philip Castle (May 18, 2020)

Philip E. Castle

Ashutosh Sabharwal

Ashutosh Sabharwal

HOUSTON, TEXAS

May 2020

ABSTRACT

Diagnosing Cervical Neoplasia in Rural Brazil:
Clinical Trial Outcomes and a Next-generation Point-of-care Imaging System
by
Brady Hunt

Cervical cancer is the most common cause of cancer mortality for women in low human-development index countries. Vaccination to prevent infection with human papillomavirus is a promising long-term solution; however, uptake of the vaccine in many low-income countries is slow. Additionally, current effective methods for screening, diagnosis and treatment of pre-cancerous cervical lesions are too costly to scale in resource-constrained healthcare settings. Even in countries where screening programs exist, loss to follow-up is very common, leading to late-stage diagnosis and poor survival rates. Developing new strategies in which patients could undergo screening, diagnostic triage, and treatment in a single visit has the potential to reduce losses to follow-up and improve access to preventative treatment, especially in medically underserved areas. For such an approach to be effective, accurate, affordable, and automated point-of-care diagnostics for cervical cancer are needed. Furthermore, rigorous field assessment of such methods in low-resource settings is greatly needed.

This work contains a series of clinical studies conducted at a Brazilian cancer hospital to validate a low-cost, portable, high-resolution microendoscope (HRME) for real-time diagnostic evaluation of cervical lesions. Findings from these clinical studies led me to explore two approaches to further improve HRME: 1) improved image

analysis algorithms and 2) improved optical instrumentation. In total, five studies are presented: three clinical studies and two technical studies.

The first clinical study evaluated the feasibility of a mobile intervention strategy utilizing the HRME to perform diagnostic follow-up of patients in rural communities. A van was adapted into a mobile diagnostic unit that could travel to the patient communities, allowing them to undergo a diagnostic examination by colposcopy and HRME close to home. A cluster-randomized study comparing diagnostic completion rates of screen positive women referred to the mobile diagnostic van versus traveling to the central hospital (current standard) was conducted. A 37% relative increase in diagnostic completion rate was observed for patients seen at the mobile diagnostic van compared to those seen at the central hospital (mobile van: 87%, central hospital: 64%, $p < 0.001$).

In the second clinical study, a retrospective case-control analysis was performed to evaluate whether or not exposure to proflavine, the contrast agent used to perform HRME imaging, was associated with progression of cervical dysplasia. There were no significant differences in disease progression between the proflavine exposed and control groups, suggesting that exposure to dilute proflavine does not increase the risk of cervical precancer and cancer progression.

The culmination of these clinical studies was the CLARA study—a prospective evaluation of the diagnostic performance of the HRME compared to expert colposcopy with cervical biopsy in 1,600 screen positive women. In this large-scale, prospective evaluation, the HRME automated image analysis algorithm was found to have 95.6% sensitivity (95%CI: 92.8–97.4%) and 56.6% specificity (95%CI: 53.7–59.4%) for detection of high-grade cervical abnormalities. Diagnostic performance was compared to the

standard-of-care method, colposcopy guided biopsy, which had 96.2% sensitivity (HRME vs colposcopy, McNemar's test: $p=0.81$) and 58.7% specificity ($p=0.18$). A stratified analysis of HRME receiver operator characteristic curves by tissue type demonstrated that benign columnar and metaplasia tissues are challenging to classify correctly using current approaches to HRME image analysis.

In the next study, the library of HRME images acquired in the aforementioned clinical studies was utilized to develop improved diagnostic classifiers using machine learning. The image dataset organized consists of 2,130 *in vivo* microscopy cervical images with corresponding biopsy results. A fully data-driven approach was undertaken using a light-weight convolutional neural network (ResNet). The optimized ResNet classifier improved area under the receiver operator characteristic curve for detection of high-grade cervical dysplasia by 6.6% (AUC 0.81 vs 0.76, DeLong's test: $p=0.02$). This algorithm approach was deployed on a consumer grade laptop with GPU acceleration, and reduced single image classification speed by 20-fold (from 1 second per image down to less than 50 milliseconds per image). Although diagnostic performance improvements of the ResNet classifier were modest, this study establishes an organized dataset with baseline performance metrics which can be utilized to further improve HRME image classifiers using machine learning.

In the final study, I present a next-generation HRME system which is capable of high-speed imaging and surveillance of larger tissue areas. This system is called the *FastHRME*, where “Fast” stands for fast acquisition with stable translation. The *FastHRME* incorporates a high-frame rate CMOS sensor, increasing acquisition frame rate from 18 frames per second up to 200 frames per second, and a 3D printed optical fiber holder for stable freehand probe movement. *FastHRME* supports probe translation

up to 15 millimeters per second while maintaining high image quality and accurate video mosaicking. The utility of these design improvements was demonstrated using optical imaging targets as well as through *in vivo* imaging experiments of the oral mucosa of healthy volunteers. *In vivo* mosaics presented contain clearly visible nuclei structures and tissue boundaries with total areas that are 20 to 40 times larger than a single HRME image field-of-view.

Taken together, the hardware and software improvements presented in these technical works can potentially enable more rapid and accurate diagnosis of cervical dysplasia at the point of care. The body of clinical and technical studies presented establish strong evidence for the potential of HRME to improve cervical cancer prevention in low-resource healthcare settings.

ACKNOWLEDGEMENTS

When I reflect on the past five years, I feel that my graduate school experience has been a very unlikely journey. A journey which not only took me across the globe numerous times, but also through some of the tallest peaks and lowest valleys of life experience. This experience has been enriched by so many wonderful individuals for whom I am deeply grateful.

A tremendous thank you to my advisor Rebecca Richards-Kortum. The discipline and work ethic you exhibit in every aspect of your life has inspired me to follow your lead. Thank you for inviting me to visit Barretos, Brazil shortly after joining your lab and for entrusting me to help carry out the clinical studies there. Thank you for being so supportive of Bailee and I through three pregnancies and losing a child. I can't imagine a better advisor to guide me through the adversity I've faced during graduate school. You did warn me at the outset that it would be a messy and difficult journey, but that in the end it would be worth it. As usual, you were right. Thank you.

I also want to thank the other members of my committee Kathleen Schmeler, Philip Castle, Ashutosh Sabharwal, and Tomasz Tkaczyk for mentoring me as a scientist and engineer. A special thanks to Kathleen Schmeler and Phil Castle for the time you have so generously invested in me. Kathleen, you have been a vital source of clinical feedback for the technologies I have worked on during this thesis. I will never forget my visits to Mozambique with you and your team. Phil, your timely and candid feedback have been instrumental in my development as a scientist.

Thanks to the amazing team of Brazilian clinical collaborators who made this work possible including (but certainly not limited to): José Humberto Tavares Guerreiro

Fregnani, Júlio César Possati Resende, Marcio Antoniazzi, Bruno de Oliveira Fonseca, Graziela de Macedo Matsushita, Naitielle de Paula Pantano, Lívia Loamí Ruyz Jorge de Paula, Viviane Andrade, and Karen Cristina Borba Souza. Working with you has forever instilled in me a love for the Brazilian people, food, and culture. Muito obrigado!

To my friends and colleagues in the RRK lab, both past and present, thank you for making the lab such a great place to work. The numerous meals you provided my family during newborn stages and the generous gifts you have organized have been deeply appreciated. Several of you have given generously of your time in mentoring and encouraging me during my time in the lab. These individuals include Meaghan Bond, Jennifer Burnett, Yubo Tang, and Jennifer Carns. I also want to make particular mention and thank those who helped me complete the research presented in this thesis including Richard Schwarz, Jennifer Carns, Katelin Cherry, Sonia Parra, Jackson Coole, Alex Kortum, David Brenes, Ruchika Mitbander, and Imran Vohra. A special thanks goes to Jackson for spending so many hours in the lab prototyping with me, and for not letting me give up on the *FastHRME*.

I would like to say a few extra words about my friend and mentor Richard Schwarz. Richard, from the time I joined the lab, you have been there for me every step of the way. I admire the meticulous and thoughtful way you go about your work. I have often observed you working tirelessly behind the scenes to move projects forward without seeking public praise. I feel so blessed to have worked so closely with you, and I cannot thank you enough for the contributions you made to this thesis.

Lastly, I want offer my greatest thanks to my family for being a constant source of encouragement and happiness in my life. Dad, thank you for the sacrifices you have

made in your life to provide ample opportunities for me in mine. I have always wanted to do everything I can to honor your sacrifices and make you proud. Mom, thank you for cultivating my mind throughout life through music, laughter, and play. Thanks for unconditionally loving me and instilling a love of others in me. To my two-year old Eliza, thanks for always reminding me how very far artificial intelligence has to go to catch up to a toddler. To my three-month old Hugh, thanks for giving me a reason to be done with this thesis sooner rather than later. To my high school sweetheart and love of my life, Bailee, thank you for being strong when I was weak, adventurous when I was boring, brave when I was fearful, hopeful when I was doubtful. You keep me balanced in life and always bring out the very best in me. No matter where our journey takes us next, I know we will make it, together.

I will be your light in the wilderness;

and I will prepare the way before you...

and ye shall know that it is by me that ye are led.

– 1 Nephi 17

In Loving Memory of My Daughter

Claire Elizabeth Hunt

CONTENTS

Abstract.....	i
Acknowledgements	v
Contents.....	viii
Nomenclature	xii
1. Introduction	1
1.1 Objective and specific aims.....	1
1.2 Overview	2
2. Background	4
2.1 Motivation and significance	4
2.2 Cervical intraepithelial neoplasia and malignant progression	5
2.3 Current and emerging methods for screening, diagnosis, and treatment.....	5
2.3.1 In vivo microscopy.....	8
2.3.2 Automated visual evaluation	9
2.4 Machine learning and convolutional neural networks	11
3. Diagnosing Cervical Neoplasia in Rural Brazil Using a Mobile Van Equipped with <i>In Vivo</i> Microscopy: A Cluster-Randomized Community Trial	14
3.1 Abstract.....	14
3.2 Introduction	15
3.3 Materials and methods.....	17
3.3.1 Study design	17
3.3.2 Participants.....	17
3.3.3 Diagnostic setting.....	19
3.3.4 In vivo microscopy.....	20
3.3.5 Diagnostic procedures.....	21
3.3.6 Histopathology	22
3.3.7 Treatment	23
3.3.8 Statistical analyses.....	24
3.4 Results.....	25
3.5 Discussion	32
4. Is Proflavine Exposure Associated with Disease Progression in Women with Cervical Dysplasia? A Brief Report.....	37
4.1 Abstract.....	37
4.2 Introduction	37
4.3 Materials and methods.....	39

4.3.1 Study participants	39
4.3.2 Follow-up and treatment	40
4.3.3 Statistical analysis	41
4.4 Results.....	43
4.4.1 Baseline patient characteristics.....	43
4.4.2 Cervical dysplasia progression	44
4.5 Discussion	48
5. Cervical Lesion Assessment using Real-time microendoscopy image Analysis: The CLARA Study	50
5.1 Abstract.....	50
5.2 Introduction	51
5.3 Materials and methods.....	52
5.3.1 Study design and participants.....	52
5.3.2 High-resolution microendoscope	53
5.3.3 Diagnostic exam procedure	54
5.3.4 Histopathology & follow-up	56
5.3.5 Statistical analyses.....	57
5.4 Results.....	59
5.5 Discussion	67
6. Benchmarking Performance of Convolutional Neural Networks for Diagnostic Classification of Cervical Dysplasia Using The CLARA Study Dataset.....	71
6.1 Overview	71
6.2 Materials and methods.....	71
6.2.1 Image dataset.....	72
6.2.2 Transfer learning using CNNs	74
6.2.3 Data preprocessing and augmentation transforms.....	76
6.2.4 Performance metrics and model selection.....	77
6.2.5 Deployment and speed benchmarks	78
6.2.6 Source code availability.....	79
6.3 Results.....	79
6.4 Discussion	84
7. FastHRME: A High Frame Rate Microendoscope Capable of Robust Freehand Mosaicking	86
7.1 Abstract.....	86
7.2 Introduction	86
7.3 Materials and methods.....	88
7.3.1 Optical assembly	88
7.3.2 Optical fiber holder	90
7.3.3 Optical target imaging.....	90

7.3.4 In vivo imaging	91
7.3.5 Video processing and mosaic construction	91
7.3.5 Quantitative metrics for assessing video mosaic quality	93
7.3.7 Graphical visualizations.....	94
7.4 Results.....	94
7.5 Discussion	102
8. Conclusion	105
8.1 Summary of results.....	105
8.2 Future work	107
9. References	109

NOMENCLATURE

AIS	Adenocarcinoma <i>in situ</i>
CIN	Cervical intraepithelial neoplasia
CMOS	Complementary metal–oxide–semiconductor
CNN	Convolutional neural network
ECC	Endocervical curettage
FOV	Field-of-view
HDI	Human development index
HPV	Human papillomavirus
HRME	High-resolution microendoscopy / microendoscope
LEEP	Loop electrosurgical excision procedure
LMICs	Low- and middle-income countries
POC	Point of care
VIA	Visual inspection with acetic acid
WHO	World Health Organization

1. INTRODUCTION

1.1 Objective and specific aims

The objective of this research is to evaluate and then improve a low-cost high-resolution microendoscope system for diagnosing cervical dysplasia in rural settings. The specific aims are as follows:

Specific Aim 1: Complete the first clinical evaluation of a mobile diagnostic strategy for cervical cancer detection using the HRME in order to: A) demonstrate the feasibility of a mobile van equipped with immediate POC diagnostic capabilities; B) compare the proportion of patients who complete diagnostic evaluation in a mobile van to the standard of care; and C) assess the safety of HRME imaging by following patients and assessing rates of disease progression for 18-24 months compared to an age-matched group of controls.

Specific Aim 2: Carry out a large prospective evaluation of HRME imaging technology to evaluate diagnostic performance in a clinical trial of 1,600 patients. Analyze resulting data to understand the conditions that give rise to incorrect results with HRME imaging and suggest hypotheses to improve performance.

Specific Aim 3: Evaluate the use of an existing convolutional neural network to analyze HRME images to diagnose cervical precancer using clinical trial data obtained in Aims 1 and 2. Compare diagnostic accuracy to image segmentation strategy used in Aim 2.

Specific Aim 4: Based on results from studies in Aims 2 and 3, develop a next-generation HRME that incorporates improved optical hardware and high-frame rate data collection in order to improve diagnostic performance and ensure good spatial correlation between the location imaged using the HRME and the site of the subsequent tissue biopsy.

1.2 Overview

This dissertation is organized as follows:

Chapter 1 lists the specific aims of the research and provides a brief overview of the contents of each chapter.

Chapter 2 provides background information on cervical cancer, standard methods for prevention and their limitations, a review of emerging optical diagnostic techniques, and a brief introduction to machine learning techniques recently being applied to cervical cancer diagnosis.

Chapter 3 presents the results of a clinical study in which a mobile intervention strategy using HRME was compared to the standard of care approach at a central hospital. Results of this trial demonstrated improved diagnostic completion rates for the mobile intervention strategy utilizing HRME as well as preliminary diagnostic performance results of the HRME. These promising results motivated a larger scale clinical trial in Brazil (described in **Chapter 5**). The contents of this chapter were previously published in *Cancer Prevention Research*.

Chapter 4 describes a retrospective analysis to observe the clinical follow-up of Brazilian patients after their exposure to proflavine during HRME imaging. This study was performed to address concerns related to proflavine use prior to conducting a large-scale clinical trial using HRME. The primary finding was that there were no significant differences in disease progression between the proflavine exposed and control groups. The contents of this chapter were previously published in *Photochemistry and Photobiology*.

Chapter 5 reports the results of a prospective evaluation of the HRME diagnostic algorithm described in **Chapter 3**. Analysis of this dataset established more precise estimation of HRME diagnostic performance. A manuscript describing this study is in the final stages of preparation. The dataset acquired during this study forms the basis for the work in image analysis algorithm development described in **Chapter 6**.

Chapter 6 presents an image analysis study performed to develop improved diagnostic classifiers for HRME. Diagnostic performance using a convolutional neural network approach is evaluated. This entirely data driven approach is compared to the previously developed algorithm based on morphological image processing of nuclei shapes within HRME images. Dramatically reduced classification speed is demonstrated using a consumer grade laptop comparable to existing HRME instrumentation.

Chapter 7 describes the design and evaluation of an improved microendoscope system, called the *FastHRME*, which utilizes a high-frame rate CMOS sensor and a 3D printed optical fiber holder. The system achieves excellent image quality at low exposures by minimizing motion blur and improving translational stability when moving the optical probe along tissue. A manuscript describing this system is in the final stages of preparation.

Chapter 8 summarizes the conclusions of this research and discusses future directions to continue the work presented.

2. BACKGROUND

2.1 Motivation and significance

Cervical cancer is the fourth most common cause of cancer incidence and mortality globally, with over a half-million new cases diagnosed and a quarter-million deaths annually.¹ The cause of cervical cancer is well understood, and effective preventative treatments have led to a significant reduction in cervical cancer mortality in high human development index (HDI) countries over the past several decades.² However, existing cervical cancer prevention methods have not been able to make a significant impact in low HDI countries, where over 85% of new cases and deaths occur today.³ The following section provides background information regarding current developments in the field of cervical cancer prevention.

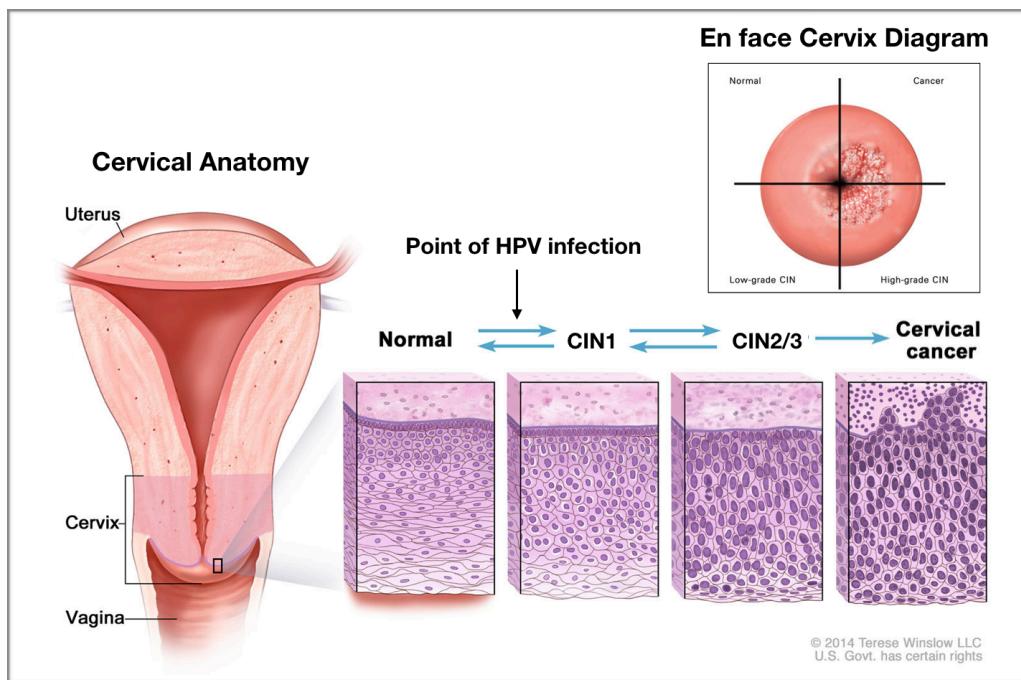


Figure 2-1 Cellular changes associated with HPV induced malignant progression of cervical epithelium. Categorical stages include normal, low-grade cervical intraepithelial neoplasia (CIN 1), high-grade cervical intraepithelial neoplasia (CIN 2 and CIN 3), and cervical cancer. Reproduced with permission.

2.2 Cervical intraepithelial neoplasia and malignant progression

The cervix is the outer opening of the uterus, which is the primary female reproductive organ (**Figure 2-1**). It is located at the back of the vagina, and can be accessed using a speculum to hold open the vaginal canal. The cervix is a dynamic tissue structure composed primarily of two cell types: squamous and columnar. These two cell types come together to form a junction near the external os called the squamocolumnar junction or transformation zone. The composition and location of the transformation zone changes over time as a result of age-related hormonal changes. The majority of cervical pre-cancerous lesions originate within the transformation zone.

Cervical cancer is caused by persistent infection of a sexually transmitted virus, human papillomavirus (HPV).⁴ HPV infection starts at the outer layer of the cervical tissue called the epithelium. Over a long period of time (years to decades), HPV infected epithelial cells can be transformed and begin to grow within the epithelium. HPV does this by integrating viral DNA into host cells, causing the host to produce proteins that disrupt the normal life cycle of the cell. This disruption of the cell cycle leads to uncontrolled proliferation of the cells, pre-cancerous lesion formation, and eventual invasion of neoplastic cells throughout the organ site. In the early stages of this process, the transformed cells that form small pre-cancerous lesions are known as cervical intraepithelial neoplasia (CIN). If CIN is detected before invasion occurs, progression to cancer can be prevented by ablation or surgical removal (excision) of the outer epithelial layer of the cervix.

2.3 Current and emerging methods for screening, diagnosis, and treatment

The goal of cervical cancer prevention programs is to identify which women are at highest risk for developing cervical cancer, and to utilize preventative treatment for those women who have developed pre-cancerous lesions. For many years the Papanicolaou (Pap) smear test has served as the primary screening test, and is utilized effectively in high-high-resource settings. In recent years, HPV testing for the oncogenic types of HPV (also known as high-risk HPV) has emerged as the most effective tool for identifying women at risk for developing cervical cancer.⁵ HPV testing is more sensitive but less specific for the detection of cervical pre-cancers and cancer than the conventional Pap smear and other cytological methods. However, HPV testing cannot differentiate between the mostly benign HPV infections (~90%) that are destined to clear and the clinically important ones (~10%) that will persist and progress to pre-cancer, which if not detected and treated, can become invasive cancer.⁶ Thus, there is a significant need for point-of-care (POC) diagnostic testing to decrease overtreatment for screen-positive women.

In high-resource settings, follow-up testing for screen-positive patients is typically visual examination of the cervix by a gynecologist with special training in colposcopy (colposcopist). The colposcopist uses magnifying optics, called a colposcope, to inspect for any abnormal tissue growth in the cervix. Based on this evaluation, the colposcopist will remove a small tissue specimen from the cervix in any areas that appear suspicious. These cervical biopsies are then analyzed by a second physician specialist, called a pathologist, who is trained to assess the specimen under a microscope.

The rigorous process employed by the pathologist to diagnose the specimen is called histopathology. After physically slicing it into thin layers, mounting those layers

onto microscope slides, and adding colored dyes to the specimen the pathologist assesses the cells on the specimen slides. Pre-cancerous lesions are characterized by cells with enlarged, densely crowded, and pleomorphic nuclei. Importantly, the pathologist assesses whether or not neoplastic cells have spread beyond the basal layer of the epithelium, which is when invasive cancer ensues. Using this information, the pathologist grades the severity of the dysplasia and provides a diagnosis to guide treatment decisions for the patient.

While colposcopy and histopathology have been highly effective components of cervical cancer prevention, they are nonetheless very costly in terms of requirements for healthcare infrastructure and highly trained personnel. Moreover, it is a very slow process for the patient, who may need to revisit the clinic multiple times over a period of weeks or months. Loss of contact and failure to complete diagnostic follow-up is a major concern in low-resource environments, where patients may live several hours away from the clinic. Due to this limitation, so called “screen-and-treat” approaches to addressing cervical cancer in low-resource areas have been proposed, with the goal of getting more timely preventative treatment to women at risk.⁷⁻⁹ While “screen-and-treat” is one approach to overcoming losses to follow-up, what is truly needed are technologies to support accurate, affordable, and automated triage and treatment of screen-positive women. Such methods would facilitate more efficient cervical cancer prevention programs in which screening, diagnosis, and treatment can be accomplished in a single visit (screen-diagnose-treat). A number of potential methods are being developed and evaluated to support screen-diagnose-treat strategies. The following sections summarize two emerging diagnostic imaging methods that apply primarily to this thesis.

2.3.1 *In vivo* microscopy

Our optics research group has pioneered development and clinical evaluation of *in vivo* microscopy with automated image analysis as a new diagnostic tool to support screen-and-treat strategies. *In vivo* microscopy is performed using a low-cost fiber-optic imaging device called the high-resolution microendoscope (HRME) (**Figure 2-2A**). The HRME gives the clinician the ability to evaluate the size, shape, and distribution of cell nuclei *in vivo* at the point of care (**Figure 2-2B**). This real-time diagnostic imaging capability could potentially be used to help guide the selection of biopsy sites as well as help identify lesions that are suitable for immediate treatment, minimizing losses to follow-up. A limitation of this technique is that the field of view of the probe is small (approximately 0.5 mm^2), so acquiring data for the entire cervix is not feasible. In 2012, the first two pilot studies using the HRME to detect evaluate cervical dysplasia were reported.^{10,11} These earliest studies performed utilized traditional computer vision techniques with manual thresholding of HRME images to segment nuclei and utilized nuclear-to-cytoplasm ratio (NC ratio) as the primary metric for diagnostic prediction. In 2015, Grant reported results from the first pilot study at Barretos Cancer Hospital (BCH) in Brazil. In that study, Grant evaluated additional nuclei segmentation metrics using

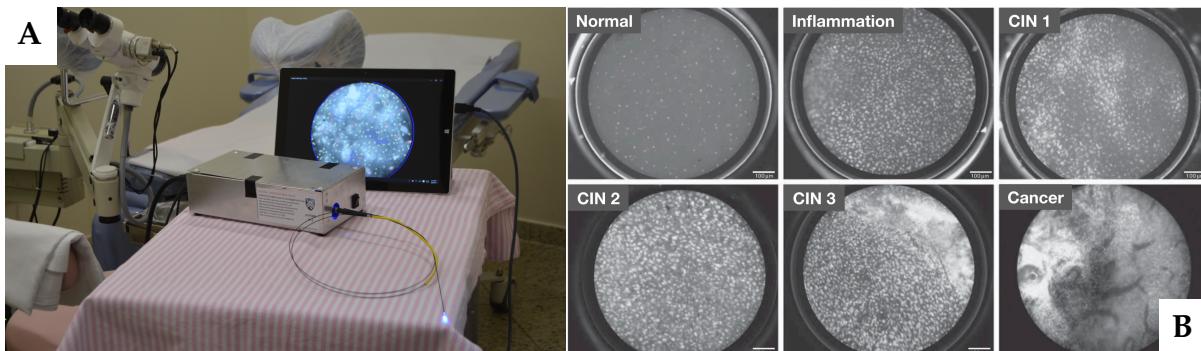


Figure 2-2 **A)** picture of the tablet-based HRME system in a cervical cancer prevention exam room in Barretos, Brazil. **B)** example HRME images of the cervix demonstrating nuclei morphological changes from malignant progression.

traditional computer vision and found that mean nuclear area and median eccentricity better discriminated high-grade abnormalities.¹²

Following these pilot studies, more automated image analysis strategies for HRME image analysis were developed for real-time diagnostic prediction and prospective algorithm evaluation. The tablet-based instrumentation and software for deployment of this approach was developed by Quang et al and first applied for real-time assessment of the esophagus.¹³ Using the same approach as Quang, Grant then evaluated additional nuclei segmentation based metrics metrics and parameter thresholds using a 228 patient dataset from Barretos Cancer Hospital (BCH).¹⁴ Grant established a metric which classifies each segmented nuclei within the image as normal or abnormal based on an area threshold of $195 \mu\text{m}^2$ and eccentricity threshold of 0.65, and then calculates the total number of abnormal nuclei per unit area. The instrumentation and algorithm parameters established by Quang and Grant have now been evaluated prospectively in more recent clinical studies conducted in Brazil, El Salvador, and the Texas-Mexico border of the United States. The largest of prospective evaluation of this instrumentation and real-time image analysis software has been in Brazil, and the results described in subsequent chapters of this thesis.

2.3.2 Automated visual evaluation

Acknowledging the high-resource requirements of colposcopy/histopathology has led to exploration of other automated visual assessment tests for use in low-resource settings. The World Health Organization (WHO) guidelines for cervical cancer prevention have recommended visual inspection with acetic acid (VIA) as a tool for training non-specialists to perform a visual evaluation of cervical lesions.¹⁵ Cervical lesions have a tendency to appear white when acetic acid applied to the cervix. After

application of acetic acid, the practitioner waits for two minutes to see if any whitening occurs. If whitening occurs, then the exam is considered positive. The primary benefit to VIA is that it can be performed fairly inexpensively and with minimal training of healthcare workers, making it feasible for low-resource settings. VIA does have relatively low specificity, which has led to discussion of finding more accurate tools that remain cost-effective and scaleable.¹⁶

More recently, automated visual evaluation of the cervix using digital colposcopy has emerged as another promising method for scalable, rapid, diagnostic evaluation of the cervix. Two promising POC colposcopy devices are the MobileODT EVA System (**Figure 2-3A**) and the Pocket Colposcope. The MobileODT EVA system is commercially available and the Pocket Colposcope is in preparation for commercial launch. Recent work with these devices has been fueled by advances in image recognition algorithms from the field of machine learning. As cervical neoplasia

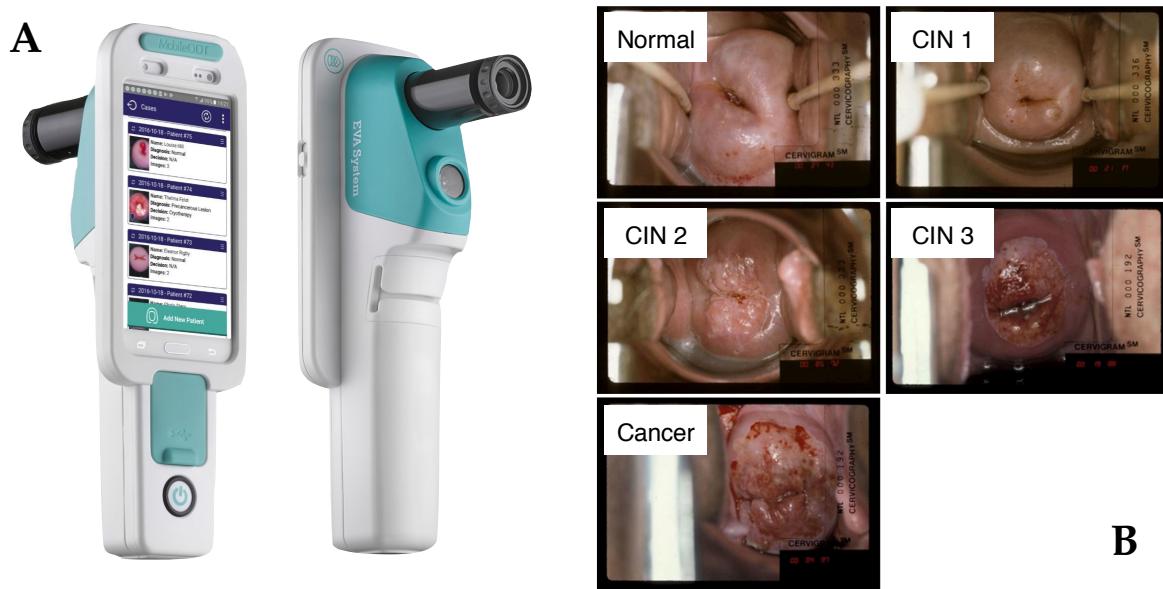


Figure 2-3 **A)** the MobileODT mobile phone colposcope system. **B)** example colposcopy images demonstrating a progression from normal to invasive cancer. Figure adapted from Hu et al.¹⁶

increases in severity, the cervix develops demarcated regions with distinct tissue texture (**Figure 2-3B**). The researchers developing the Pocket Colposcope recently reported diagnostic performance of 81% sensitivity and 79% specificity for detection of CIN 2+ on a cohort of 134 patients.¹⁷ Their automated algorithm used texture-based features and a support-vector machine classifier. Another study reported on a convolutional neural network classifier that was trained on a much larger screening population of 9,406 patients, achieving 93% sensitivity and 83% specificity (AUC=0.91).¹⁸ This latter result was published in a prominent clinical cancer journal and has given rise to discussion about primary screening using automated visual evaluation directly.

2.4 Machine learning and convolutional neural networks

Machine learning is "the scientific study of algorithms and statistical models that computer systems use to progressively improve their performance on a specific task".¹⁹ A central concept in machine learning is gradient-based optimization of predictive models. Predictive models use data as numerical inputs and compute predicted outputs based on the model parameters. In order to measure performance of a model, an error function must be defined which quantifies how well the prediction matches the desired output. Once a performance metric is defined, model parameters can be explored that will minimize the average error of the model on the dataset. High-dimensional datasets inherently require a larger number of model parameters, and can often be less intuitive to interpret. Machine learning applications often deal with very high-dimensional datasets and models.

To ensure the development of generalizable models, machine learning is commonly broken in two-phases: training and testing. First, two distinct subsets of the data are created from all available data—a training dataset and a testing dataset. The

purpose of the training phase is to find model parameters to make accurate predictions on the data, whereas the purpose of testing is to simulate a prospective evaluation of the model on future data. Gradient descent is an algorithm that iteratively updates parameters of the model based on the multivariate derivative (also known as the gradient) of the error function. The reason an iterative process is often necessary is because of high-dimensionality inputs and large numbers of training examples. Given the model inputs and desired model outputs, gradient descent guides the model parameters towards a local minimum of the error function.

Neural networks are a particular type of predictive models that have gained increasing popularity in recent years. Yann LeCun, Geoffrey Hinton, and Yoshua Bengio are three of the primary researchers credited with the development of neural networks.²⁰ Seminal work by LeCun in 1989 showed arguably the first practical application of neural networks to the problem hand-written zip code digit recognition.^{21,22} His work established a particular type of neural network model known now as a convolutional neural network (CNN). The foundational concepts of CNNs have not changed significantly since LeCun's work, but the computing power and availability of large datasets have led to a revival in application of CNNs to various tasks. This reincarnation of convolutional neural networks using neural networks with a large number of layers has been termed "deep learning".²⁰

In recent years, there has been an explosion of frameworks being developed for creating and optimizing neural networks. Many of the major technology companies have been active in this area. Two of the frontrunners are TensorFlow and PyTorch, which are open-source frameworks published by Google and Facebook respectively.^{23,24} Both frameworks enable easy construction of custom neural network models, with

efficient parallelization of CNN optimization over high-performance GPUs. These frameworks have enabled non-experts to train and deploy CNNs and have played a large role in the spread of machine learning research into many new applications, such as the domain of cervical cancer prevention.

3. DIAGNOSING CERVICAL NEOPLASIA IN RURAL BRAZIL USING A MOBILE VAN EQUIPPED WITH *IN VIVO* MICROSCOPY: A CLUSTER-RANDOMIZED COMMUNITY TRIAL

The contents of this chapter have been published in the following journal article: Hunt B, Fregnani JH, Schwarz RA, Pantano N, Tesoni S, Possati-Resende JC, Antoniazzi M, de Oliveira Fonseca B, de Macêdo Matsushita G, Scapulatempo-Neto C, Kerr L. Diagnosing cervical neoplasia in rural Brazil using a mobile van equipped with *in vivo* microscopy: A cluster-randomized community trial. *Cancer Prevention Research* 11.6 (2018): 359-370.

3.1 Abstract

Background: Cervical cancer is a leading cause of death in underserved areas of Brazil.

We compared rates of diagnostic follow-up completion for women with abnormal cervical cancer screening results attending a mobile diagnostic van in their local community versus those asked to travel to a central hospital for diagnosis. We also evaluated the diagnostic performance of *in vivo* microscopy compared to colposcopy.

Methods: This prospective randomized trial involved 200 women in southern/central Brazil with abnormal Papanicolaou tests. Participants were randomized by geographic cluster and referred for diagnostic evaluation either at a mobile van upon its scheduled visit to their local community, or at a central hospital. Participants in both arms underwent colposcopy, *in vivo* microscopy, and cervical biopsies.

Results: There was a 23% absolute and 37% relative increase in diagnostic follow-up completion rates for patients referred to the mobile van (102/117, 87%) compared with the central hospital (53/83, 64%) ($p=0.0001$; risk ratio=1.37, 95%CI=1.14-1.63). In 229 cervical sites in 144 patients, colposcopic examination identified sites diagnosed as cervical intraepithelial neoplasia grade 2 or more severe (CIN 2+; 85 sites) with a sensitivity of 94% (95%CI=87%-98%) and specificity of 50% (95%CI=42%-58%). *In vivo* microscopy with real-time automated image analysis identified CIN 2+ with a sensitivity of 92% (95%CI=84%-97%) and specificity of 48% (95%CI=40%-56%).

Conclusion: Women referred to the mobile van were more likely to complete their diagnostic follow-up compared to women who were referred to a central hospital, without compromise in clinical care. *In vivo* microscopy in a mobile van provides automated diagnostic imaging with sensitivity and specificity similar to colposcopy.

3.2 Introduction

Cervical cancer remains a major global public health problem despite effective screening with Papanicolaou (Pap) and human papillomavirus (HPV) DNA testing as well as the more recent introduction of the HPV vaccine. Over 500,000 new cases and 270,000 deaths due to cervical cancer still occur annually worldwide.⁽¹⁾ Over 85% of cases and deaths occur in women in low- and middle-income countries (LMICs), where Pap- or HPV DNA-based screening and follow-up programs are often lacking or flawed.^{2,7,25}

In Brazil, approximately 16,400 new cases of cervical cancer occur each year, representing the 3rd highest cancer incidence among Brazilian women, excluding non-melanoma skin cancer.²⁶ Incidence rates vary within Brazil with the highest rates in underserved areas located far from screening and diagnostic centers.²⁶ One strategy to reach patients in these areas, developed by Barretos Cancer Hospital in the northern part of São Paulo state, is the use of medically equipped semi-trailer trucks that travel to underserved communities to perform Pap-based cervical cancer screening and other services.²⁷⁻³³ While this is effective in terms of initial screening, patients with an abnormal Pap test result must still travel to a central hospital for diagnostic follow-up (colposcopy, biopsy, and histopathologic evaluation) – often a journey of two or more days by bus. It is estimated that up to two thirds of screen-positive women fail to return

to the central hospital for the necessary diagnostic workup and treatment, especially when they live far from the central facility.³⁰

In this study, we explored a strategy for bringing mobile diagnostic follow-up to underserved communities in Brazil. This was achieved through the use of a mobile diagnostic van equipped with a colposcope, *in vivo* microscope, cryotherapy equipment, and staffed by a colposcopist and a research nurse. The mobile diagnostic van traveled from Barretos Cancer Hospital to communities where patients with abnormal Pap results had previously been identified via mobile screening trucks. Patients received diagnostic follow-up care in the van including colposcopy, *in vivo* microscopy, and cervical biopsies. The *in vivo* microscope is an important component of this strategy and consists of a low-cost fiber-optic imaging device called the high-resolution microendoscope (HRME) that can image subcellular morphologic features in real time.^{11,12} This gives the clinician the ability to evaluate the size, shape, and distribution of cell nuclei *in situ* at the point of care – information that is typically only available weeks to months later after taking a biopsy, tissue processing, and evaluation at the central hospital. This real-time diagnostic imaging capability could potentially be used to help guide the selection of biopsy sites and reduce the number of biopsies needed for diagnosis. It could also potentially help identify lesions that are suitable for immediate treatment using cryotherapy, minimizing losses to follow-up.

The primary objective of this study was to compare rates of diagnostic follow-up completion for screen-positive women randomized to attend a mobile diagnostic van in their local community versus a central hospital for diagnostic follow-up. The secondary objective of the study was to evaluate the diagnostic performance of *in vivo* microscopy compared to colposcopy using histopathology as the gold standard.

3.3 Materials and methods

3.3.1 Study design

We conducted a non-blinded, cluster-randomized community trial to evaluate two strategies for follow-up diagnosis of cervical cancer precursor lesions in patients with an abnormal Pap test result: 1) referral of patients to a mobile diagnostic van upon its visit to their local community (experimental arm) versus 2) referral of patients to travel to a central hospital for diagnostic follow-up (standard of care arm). Subjects were randomized by geographic cluster to the experimental arm or the standard of care arm.

The study was approved by the Barretos Cancer Hospital Ethics Research Committee, the Brazilian National Ethics Research Commission (CAAE: 37774314.3.0000.5437), and the institutional review boards of Rice University (ID#653693) and the University of Texas MD Anderson Cancer Center (ID#2015-0442). Written informed consent was obtained from all participants. The protocol was registered with ClinicalTrials.gov (NCT 02494310).

3.3.2 Participants

Potential subjects were identified through a regional Pap Smear screening program operated by Barretos Cancer Hospital (BCH) via mobile screening trucks. Each city scheduled to be visited by the mobile screening trucks was eligible for cluster randomization. Randomization was done in three rounds as the schedule for the mobile screening trucks was not known in advance. In each round, a list of cities was provided by clinicians at BCH to collaborators at Rice University. Each city was assigned a number and a computer program was used generate a random number sequence for the number of cities. The first half of the cities in the sequence was assigned to the

experimental arm, whereas the second half was assigned to the standard care arm. A total of 234 geographic clusters were randomized; 79 clusters with patients having abnormal Pap results were included in the study (**Table 3-1**). Per Brazilian guidelines, women 18 years of age or older who had an abnormal screening Pap test (\geq ASC-US) were referred for colposcopy and therefore eligible for the study and invited by telephone to participate.

Invited to attend mobile van			Invited to attend central facility		
City	HDI	Distance (km)	City	HDI	Distance (km)
Água Boa	0.729	1094	Adamantina	0.790	322
Alto Taquari	0.705	695	Alta Floresta	0.714	1987
Aral Moreira	0.633	356	Altair	0.687	93
Araputanga	0.725	1532	Alto Garças	0.701	831
Bebedouro	0.780	51	Bataguassu	0.710	335
Bom Jesus do Araguaia	0.661	1341	Cáceres	0.708	1361
Brasnorte	0.696	1749	Campanha	0.709	467
Campinápolis	0.538	1096	Campo Novo do Parecis	0.734	1585
Campo Verde	0.750	1115	Cocalinho	0.660	1031
Castelo dos Sonhos (Altamira)	0.665	2059	Colina	0.757	21
Dourados	0.747	229	Cruzeiro do sul	0.713	528
Euclides da Cunha Paulista	0.704	527	Diamante do Norte	0.723	602
Gabriel Monteiro	0.763	256	Guairaca	0.693	624
General Carneiro	0.670	922	Guaraci	0.737	64
Glória d'Oeste	0.710	1514	Guiratinga	0.705	911
Glória de Dourados	0.721	281	Inajá	0.705	537
Guaíra	0.753	40	Marilena	0.681	631
Itaguajé	0.707	504	Mirassol d'Oeste	0.704	1490
Itiquira	0.693	897	Moraes De Almeida (Itaituba)	0.640	2254
Jardim Olinda	0.682	498	Noroagros (Comodoro)	0.689	1779
Jauru	0.673	1597	Nova Bandeirantes	0.650	2150
Juína	0.716	1926	Nova Lacerda	0.636	1726
Juruena	0.662	2077	Nova Xavantina	0.704	1008
Marcelândia	0.701	1883	Novo Progresso	0.673	2155
Nova Monte Verde	0.691	2140	Novo Santo Antônio	0.653	1408
Palestina	0.732	140	Novo São Joaquim	0.649	1051
Paranáita	0.672	2058	Ourinhos	0.778	375
Paranhos	0.588	467	Ponta Porã	0.701	313
Querência	0.692	1258	Porto dos Gaúchos	0.685	1788
Redenção	0.672	1670	Primavera do Leste	0.752	1107
Rosana	0.764	573	Ribeirão Cascalheira	0.670	1203
Santa Cruz do Xingu	0.684	1559	Rondonópolis	0.755	981
São Gotardo	0.736	386	São José do Povo	0.661	987
São José dos Quatro Marcos	0.719	1504	São José do Xingu	0.657	1488
Serra Nova Dourada	0.664	1349	Sete Quedas	0.614	469
Taiacú	0.710	83	Tabaporã	0.695	1872
Terra Roxa	0.749	53	Taquaral	0.759	64
Três Pontas	0.731	377	Teodoro Sampaio	0.741	478
			União do Sul	0.665	1856
			Vista Alegre do Alto	0.744	121
			Xinguara	0.646	1783
Mean	0.700	996	Mean	0.699	1020

Table 3-1 Comparison of human development index (HDI) and distance to central facility for the 79 geographic clusters.

Each geographic cluster was characterized by its distance from the central hospital (BCH) and by its human development index (HDI), which is a composite statistic of life expectancy, education, and per capita income. HDI values for each geographic cluster/municipality were obtained from an official Brazilian government website, Instituto Brasileiro de Geografia e Estatística, at:

<https://cidades.ibge.gov.br/xtras/home.php>, accessed May 2016.

3.3.3 Diagnostic setting

Patients in the standard of care arm traveled to the colposcopy clinic at BCH for the required diagnostic procedures. Patients in the experimental arm underwent diagnostic procedures in a mobile diagnostic van that traveled to their community (**Figure 3-1**). The interior of a Mercedes Sprinter 515 CD1 van was configured to include two rooms: a changing room in the rear of the vehicle, and an examination room equipped with a table for gynecologic examinations, colposcope, *in vivo* microscope, supplies, washbasin, electrical power (when plugged in to an external power source), running water, and air conditioning. All vehicle adaptions were made in the Mobile Unit Factory of BCH (Barretos Lamboo Company, Brazil).

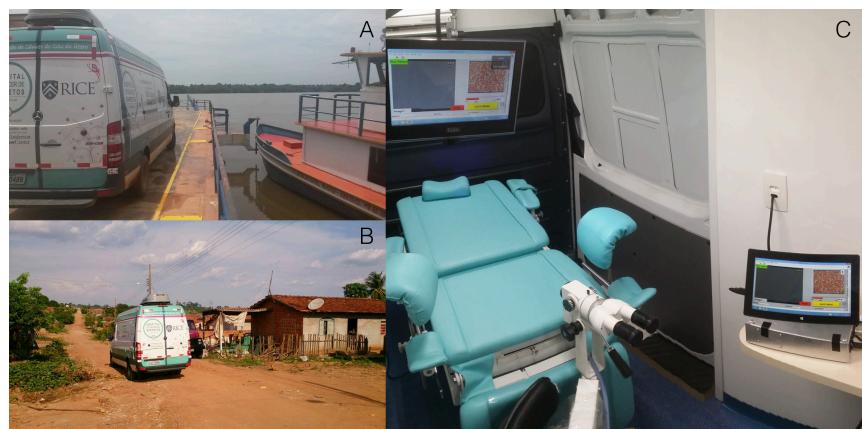


Figure 3-1 The mobile diagnostic van: **A)** In transit at the Juruena river, **B)** Arriving at a remote community in the state of Mato Grosso, **C)** View of the examination room inside the van.

3.3.4 In vivo microscopy

The HRME was developed at Rice University (Houston, Texas, USA). It is a portable, battery-powered fluorescence microscope with a flexible fiber-optic probe.^{13,34,35} The HRME has a field of view of 790 microns and transverse resolution of 4 microns; the probe is flexible but has a minimum recommended radius of curvature of 90 mm. The HRME is used in combination with the topically applied contrast agent 0.01% proflavine. During use, the clinician holds the HRME probe a few inches back from the distal tip and places the probe in contact with the tissue surface. The HRME device images the size, shape, and distribution of epithelial cell nuclei.

Upon image acquisition, the HRME software automatically processes the image, performs a quality control check, and applies a quantitative diagnostic algorithm based on morphologic features including the size, eccentricity, and crowding of cell nuclei.^{10,36-38} The automated algorithm segments individual nuclei and outlines them in the image display in yellow if they fall within a normal range for size and eccentricity as defined in the software, or in red if they exceed those limits. After segmenting the image in this manner, the quality control check is performed using two quantitative metrics: raw pixel intensity of the fiber bundle region and signal to background ratio of the segmented nuclei compared to the segmented cytoplasm areas of the image. Finally, the automated algorithm calculates a quantitative parameter defined as the number of abnormal nuclei per square millimeter of area included in the analyzed field of view. This parameter, #abnormal nuclei / mm², is used to categorize the overall image as non-neoplastic or neoplastic, with a preestablished threshold value of 120 abnormal nuclei / mm² as the cutoff. The threshold values for quality control and diagnosis were established based on data from a previous cervical dysplasia imaging study¹² and data

from the first 34 patients in this study (22 in the central facility arm and 12 in the mobile van arm). Starting with the 35th patient in this study, the algorithm, parameters, and thresholds were fixed throughout the remainder of the study. The HRME displays this quantitative result and a diagnostic prediction for the site within approximately 6-8 seconds following image acquisition.

3.3.5 Diagnostic procedures

Diagnostic procedures were carried out in an identical fashion in the central hospital and the mobile van. Procedures were performed by one of three colposcopists with similar training and clinical experience (JCPR, MA, BOF). All three colposcopists participated in both arms of the study.

An initial examination was performed using visual inspection with 5% acetic acid followed by colposcopic examination. Colposcopy images were acquired using a KLP 210 colposcope (Koloplast ci Ltda, São Paulo, Brazil) in the mobile van and CP-M1255 colposcope (DFV, Rio de Janeiro, Brazil) in the central facility. Any colposcopically abnormal areas were identified and recorded as 'low-grade' (LG), 'high-grade' (HG), or 'cancer'. By default, any areas not identified as abnormal were categorized as 'no lesion' by colposcopy. Colposcopic calls of low-grade or more abnormal were considered positive, consistent with standard of care at BCH. All colposcopy images and impressions were recorded prior to *in vivo* microscopy. Colposcopy images in addition to colposcopic impressions were entered into the study database (REDCap), but only the impressions were used in the diagnostic performance analysis presented.

After colposcopy, proflavine at a concentration of 0.01% was then applied to the cervix and *in vivo* microscopy was performed using the HRME device. HRME images

were acquired from colposcopically abnormal cervical sites (with HRME probe placement at these sites guided by colposcopy). HRME images were also acquired from a minimum of one site per quadrant, even if no colposcopically abnormal lesion was present (with HRME probe placement at these sites guided by visual examination). The colposcopist was tasked with collecting a single HRME image from each site, but had discretion to collect more than one image if desired. If more than one image was collected at a single site, the image with the highest (most abnormal) quantitative score according to the automated algorithm was used. The diagnostic prediction generated by the automated HRME software was noted for each site. Sites identified as abnormal by colposcopy or by HRME imaging were biopsied. If no sites were identified as abnormal by either method, a single biopsy was taken from a clinically normal site randomly selected by the clinician. Endocervical curettage (ECC) was performed in a small number of patients when clinically indicated (mobile van: 9 patients; central facility: 3 patients). Because HRME is not capable of imaging the endocervix, pathology resulting from ECC does not correspond to cervical sites where images were acquired. Therefore, ECC pathology was not used to evaluate diagnostic performance.

3.3.6 Histopathology

Biopsy specimens underwent standard processing with hematoxylin and eosin staining as per standard procedure. Two expert pathologists (GMM, CSN or LK) who were blinded to all study results reviewed the histologic slides and classified each site as benign (including normal, atrophy, and inflammation), cervical intraepithelial neoplasia grade 1 (CIN 1), grade 2 (CIN 2), grade 3 (CIN 3), adenocarcinoma *in situ* (AIS), or invasive carcinoma, according to standard criteria defined by the World Health Organization.³⁹ Discrepant results were resolved by consensus review.

3.3.7 Treatment

Patients in the standard of care arm with a histopathology result of CIN 2 or more severe (CIN 2+) were referred within the central hospital facility for loop electrosurgical excision procedure (LEEP), cold knife conization (CKC) or cancer management based on the results. Patients in the standard of care arm with histopathology results of <CIN 2 were recommended to have follow-up in 6 and / or 12 months per local standard of care. Cryotherapy was not offered to subjects in the standard of care arm.

For patients in the experimental arm with a lesion noted by colposcopy, the option of immediate treatment using cryotherapy in the mobile diagnostic van was available, provided the lesion covered less than 75% of the ectocervix, could be fully covered with the cryotherapy tip, did not extend into the endocervical canal, and was not suggestive of cancer.⁴⁰ These are standard local eligibility criteria for cryotherapy. The assessment of eligibility for cryotherapy was made on the basis of visual inspection and colposcopic imaging. Women with a lesion noted by colposcopy who were ineligible for cryotherapy or who chose not to get cryotherapy were referred to the central hospital facility for follow-up and / or treatment as per standard of care. Women without a lesion noted by colposcopy were recommended to have follow-up. In any case where subsequent histopathology indicated CIN 2+, the patient was referred to the central hospital facility for treatment as per standard of care.

Patients in the experimental arm who received cryotherapy but whose histopathology results subsequently indicated AIS or invasive carcinoma were recalled to a central hospital for immediate treatment. Patients who received cryotherapy and

whose histopathology results subsequently indicated CIN 2 or CIN 3 were scheduled for follow-up at the central hospital in 6 months per local standard of care.

3.3.8 Statistical analyses

Sample size calculations for the study design were performed based on a cluster randomization power analysis using the two-sided score test.⁴¹ Sampling 20 clusters with five subjects each in both arms provided 81% power ($p=0.05$) to detect a 30% absolute increase in diagnostic follow-up completion rates for screen-positive women in the experimental arm, assuming that only 40% of the women with an abnormal Pap test would return in the standard of care arm. All data were collected and managed using the REDCap (Research Electronic Data Capture) platform.⁴² Statistical significance testing was performed using R 3.2.2.⁴³

The primary endpoint was the rate of diagnostic follow-up completion for women in each arm of the study. Diagnostic completion rates were compared using Fisher's exact test. The relative difference in completion rates was quantified using a risk ratio. The risk ratio and the 95% confidence interval (95%CI) of completing diagnostic follow-up were calculated using unconditional maximum likelihood estimation (Wald).⁴⁴

The secondary endpoint was diagnostic performance (sensitivity and specificity) of colposcopy and *in vivo* microscopy for detection of CIN 2+. Both low-grade or more abnormal (LG+) and high-grade or more abnormal (HG+) were used as thresholds for positivity to evaluate the diagnostic performance of colposcopy. Sensitivity and specificity with 95% binomial exact confidence intervals were calculated on both a per-site basis and a per-patient basis, with the most abnormal colposcopy result, HRME result, and histopathology result used in the per-patient analysis. Significance testing

for the two diagnostic methods for sensitivity and specificity was performed using McNemar's test, which is a statistical test used to determine whether there are differences in a dichotomous dependent variable between two related groups.⁴⁵

3.4 Results

The study schema is shown in **Figure 3-2**. Following cluster randomization, 83 patients from 41 geographic clusters were assigned to the standard of care arm (central hospital) and 117 patients from 38 geographic clusters were assigned to the experimental arm (mobile diagnostic van). There were no significant differences between the geographic clusters in the standard of care and experimental arms with regards to mean human development index (0.699 and 0.700 respectively) and mean distance to the central hospital facility (1020 km and 996 km respectively) (**Table 3-1**).

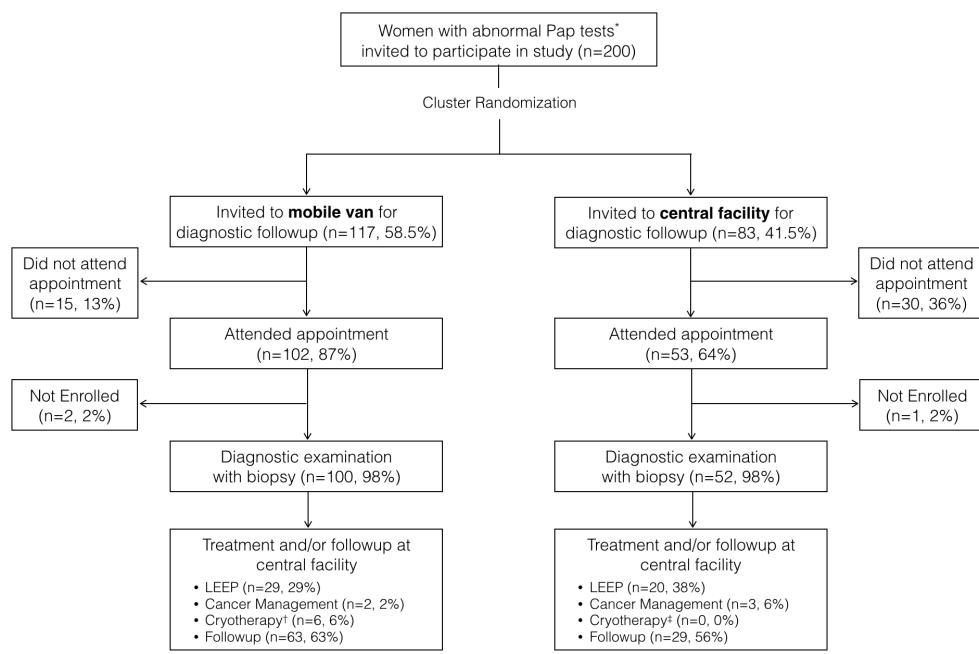


Figure 3-2 Comparison of the diagnostic follow-up rates between the mobile van and central hospital facility. Treatment and/or follow-up information reflects management recommendations, which in some cases may differ from treatment and/or follow-up actually received (except in the case of cryotherapy in the mobile van arm, which was performed immediately).

	Number of patients invited	Number (%) of patients attended	p-value*	Risk ratio	95% Confidence interval
Mobile diagnostic van	117	102 (87%)			
Central facility	83	53 (64%)	0.0001	1.37	1.14 – 1.63

* p-value calculated using Fischer's exact test

Table 3-2 Diagnostic completion rates for patients who received diagnostic follow up at the central hospital facility compared to the diagnostic mobile van.

The primary outcome results are shown in **Table 3-2**: there was a 23% absolute and 37% relative increase in diagnostic completion rates in the mobile van (102 of 117 patients, 87%) compared with the central hospital facility (53 of 83 patients, 64%) ($p=0.0001$; risk ratio=1.37, 95%CI=1.14-1.63). As the risk ratio of 1.37 indicates, the attendance rate for women in the mobile van arm was 37% higher relative to the attendance rate for women in the central hospital arm. Of the 155 patients who presented for diagnostic follow-up, 152 underwent both colposcopy and *in vivo* microscopy (100 patients in the mobile van and 52 patients in the central hospital facility). Three patients who presented for diagnostic follow-up (two in the mobile van and one at the central facility) did not undergo these procedures: one underwent LEEP treatment at a different facility, one had excessive bleeding preventing colposcopy, and one had a scheduling error. The presenting Pap findings, final biopsy diagnoses, and referral to treatment for these 152 patients are summarized in **Table 3-3**.

	Mobile van	Central facility	
No. patients	100	52	
Presenting Pap diagnosis – no. patients (%)			p-value*
ASC-US	36 (36%)	13 (25%)	0.18
ASC-H	23 (23%)	13 (25%)	
LSIL	26 (26%)	10 (19%)	
HSIL	12 (12%)	13 (25%)	
AGC	3 (3%)	3 (6%)	
Final biopsy diagnosis – no. patients (%)			p-value
Benign	40 (40%)	17 (33%)	0.27
CIN1	30 (30%)	13 (25%)	
CIN2	14 (14%)	7 (13%)	
CIN3	11 (11%)	14 (27%)	
Adenocarcinoma in-situ	1 (1%)	0 (0%)	
Invasive carcinoma	2 (2%)	1 (2%)	
Indeterminate	2 (2%)	0 (0%)	
Referral to treatment – no. patients (%)			p-value
Follow-up	63 (63%)	29 (56%)	0.11
Cryotherapy (only offered in mobile van arm)	6 (6%)	0 (0%)	
LEEP	29 (29%)	20 (38%)	
Cancer management	2 (2%)	3 (6%)	

* p-values calculated using Fischer's exact test. p-values compare the populations in the two arms of the study in terms of presenting Pap diagnosis, final biopsy diagnosis, and referral to treatment. No statistically significant difference was observed between the populations in terms of those categories ($p>0.05$).

Treatment and/or follow-up information reflects management recommendations, which in some cases may differ from treatment and/or follow-up actually received (except in the case of cryotherapy in the mobile van arm, which was performed immediately).

ASC-US (atypical squamous cells of undetermined significance), ASC-H (atypical squamous cells, cannot exclude HSIL), LSIL (low grade squamous intraepithelial lesion), HSIL (high grade squamous intraepithelial lesion), AGC (atypical glandular cells), CIN (cervical intraepithelial neoplasia), LEEP (loop electrosurgical excision procedure)

Table 3-3 Screening Pap results, final biopsy diagnosis, and referral to treatment for the 152 patients who completed diagnostic follow-up.

A total of 625 cervical sites in these 152 patients were examined using colposcopy and HRME imaging and the results compared to the gold standard of histopathologic diagnosis. We excluded 396 of the cervical sites from the diagnostic performance analyses due to 1) normal findings with no biopsy taken ($n=350, 56\%$), 2) HRME images failed an automated quality control check ($n=39, 6\%$), and 3) indeterminate histopathology ($n=7, 1\%$). The majority of images that failed the automated quality control check were those with very low overall intensity, which can arise from poor

contact between the fiber probe and cervical tissue. Of the seven cases with indeterminate histopathology, in three cases both pathologists' individual reads indicated an indeterminate result or insufficient tissue for analysis. In four cases, the pathologists' individual reads did not agree and they were unable to reach a clear consensus result upon further review. As a result of the exclusions applied to cervical sites in the dataset, a total of 8 patients were excluded from the per-patient analysis (3 with overall indeterminate pathology and 5 with no HRME images passing the automated quality control check).

The data from the remaining 229 cervical sites in 144 patients were included in the colposcopy and HRME performance analyses. Of these 229 cervical sites, 187 sites had a single HRME image collected and 42 sites had two HRME images collected. At sites where two HRME images were collected, the image with the highest (most abnormal) quantitative score was used. A histopathologic diagnosis of CIN 2+ was diagnosed in 85 of the 229 cervical biopsy sites (37%) and in 50 of the 144 patients (35%). On a per-site basis, the distribution of histopathologic diagnoses was as follows: 87 benign; 57 CIN 1; 39 CIN 2; 39 CIN 3; 7 cancer. On a per-patient basis, the distribution of histopathologic diagnoses was as follows: 51 benign; 43 CIN 1; 21 CIN 2; 25 CIN 3; 4 cancer.

Examples of colposcopic and HRME images for sites in three different patients are shown in **Figure 3-3**. The CIN 1 cervical lesion in **Figure 3-3A** was identified as negative by both colposcopy and HRME imaging. The CIN 3 cervical lesion in **Figure 3-3B** was identified as a high-grade lesion by both colposcopy and HRME imaging. However, the cervical lesion in **Figure 3-3C** showed inflammation and was incorrectly

identified as positive by both colposcopy and HRME, illustrating the possibility of a false positive result due to the confounding effects of inflammation.

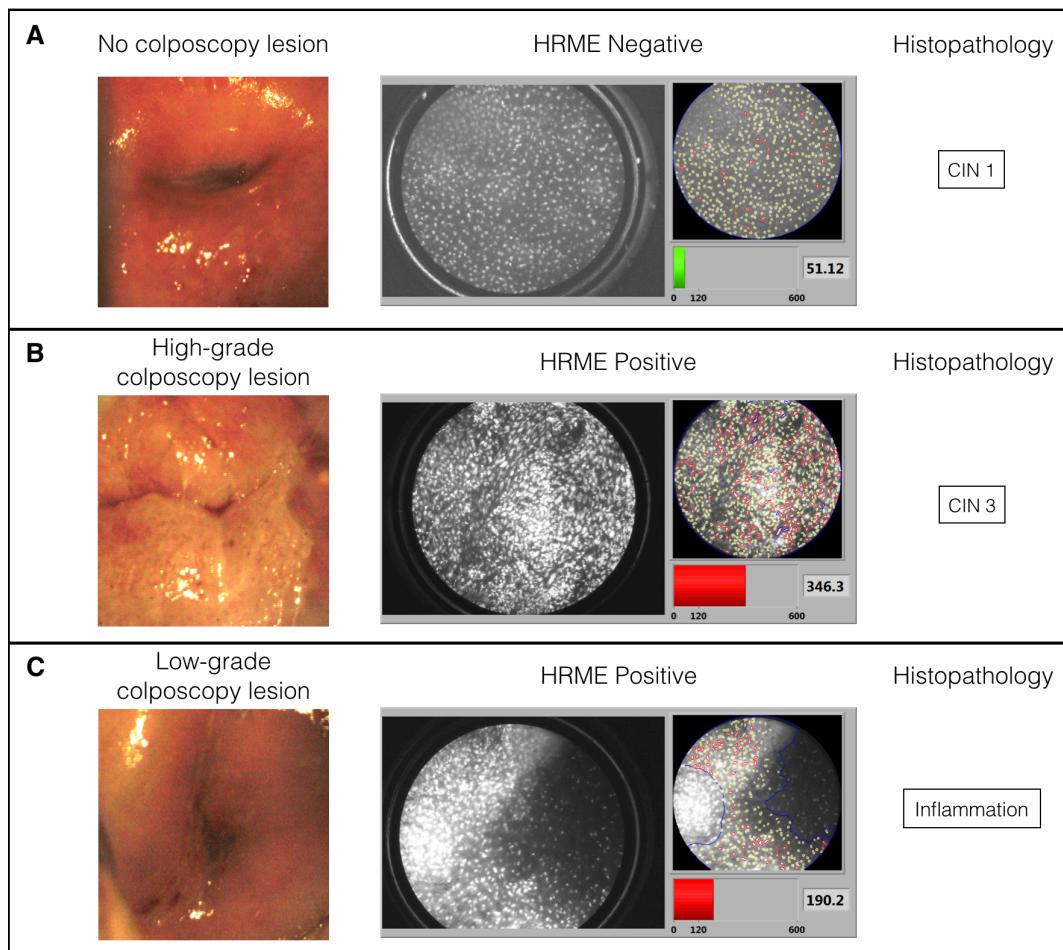


Figure 3-3 Colposcopic and HRME images for three example patients. The histological result is shown for reference. The HRME panels show a raw image of the nuclear morphology (left) and the automated analysis with quantitative interpretation (right). **A)** Both methods correctly identify the cervical site as benign. **B)** Both methods correctly identify the cervical site as a pre-cancerous lesion. **C)** Both methods incorrectly indicated a pre-cancerous lesion with the biopsy result showing inflammation. In the overlay, nuclei within normal limits for size and eccentricity are outlined in yellow; nuclei exceeding those limits are outlined in red. The parameter #abnormal nuclei/mm² is displayed at bottom right and in the bar graph relative to the threshold value (120). The bar is green if <120 (predicted non-neoplastic) or red if >120 (predicted neoplastic).

Inflammation accounted for a significant portion of the false positives for both colposcopy and HRME. A higher frequency of inflammation was observed among patients who were false positive by HRME than those who were false positive by

colposcopy, though not statistically significant (per patient: 17/33 (52%) inflammation among HRME false positives; 14/38 (37%) inflammation among colposcopy false positives; $p=0.24$; per site: 48/75 (64%) among HRME false positives; 35/72 (49%) among colposcopy false positives; $p=0.07$).

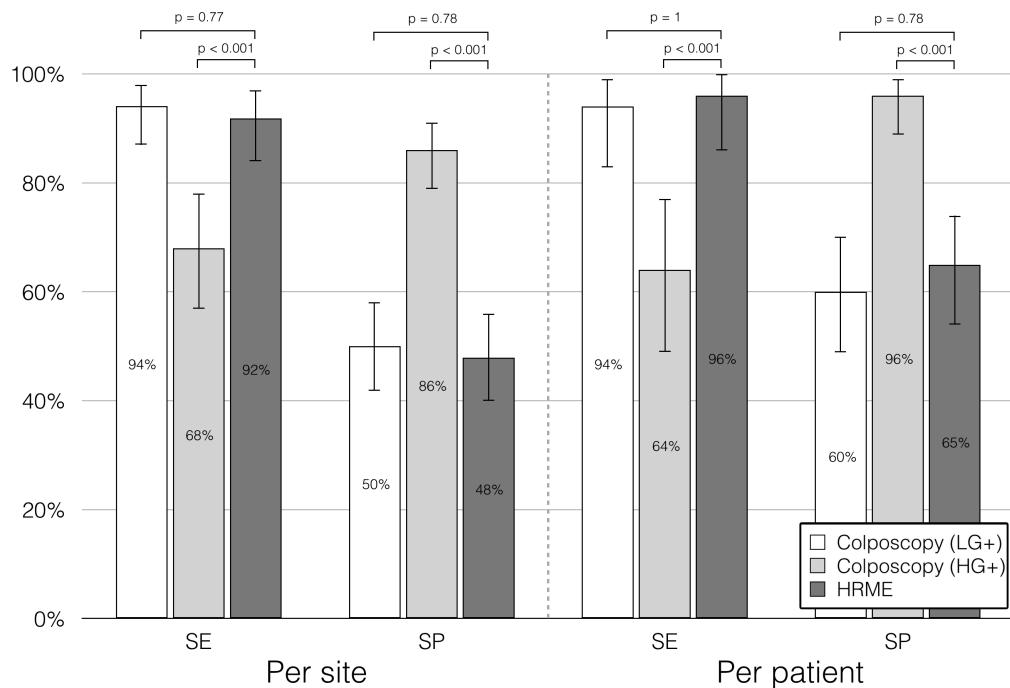


Figure 3-4 Sensitivity (SE) and specificity (SP) of colposcopy and four-quadrant HRME for detecting CIN 2+ per biopsy site and per patient. Colposcopy LG+ and HG+ indicate performance of colposcopy where low-grade and high-grade were used as the threshold for positivity, respectively. Error bars represent binomial exact 95% confidence intervals. Significance testing was performed using McNemar's test.

Figure 3-4 shows the secondary outcome results of the study: the sensitivity and specificity of colposcopy and *in vivo* microscopy for identification of CIN 2+ with respect to histopathology. The sensitivity and specificity (with 95% binomial exact confidence intervals in parentheses) were as follows: on a per cervical lesion site (per-site) basis, low-grade colposcopy (threshold of low-grade or more abnormal) had a sensitivity of 94% (87%-98%) and a specificity of 50% (42%-58%), high-grade colposcopy (threshold of high-grade or more abnormal) had a sensitivity of 68% (57%-78%) and a

specificity of 86% (79%-91%), and *in vivo* microscopy had a sensitivity of 92% (84%-97%) and specificity of 48% (40%-56%); on a per-patient basis, low-grade colposcopy had a sensitivity of 94% (83%-99%) and specificity of 60% (49%-70%), high-grade colposcopy had a sensitivity of 64% (49%-77%) and specificity of 96% (89%-99%), and *in vivo* microscopy had a sensitivity of 96% (86%-100%) and specificity of 65% (54%-74%). No significant statistical difference ($p>0.05$) was found between low-grade colposcopy and *in vivo* microscopy in terms of sensitivity and specificity on a per-site or per-patient basis. High-grade colposcopy was found to be significantly more specific than *in vivo* microscopy ($p<0.001$); however, *in vivo* microscopy was significantly more sensitive ($p<0.001$), indicating a trade-off between sensitivity and specificity for high-grade vs low-grade colposcopy.

Table 3-4 shows the detailed results of *in vivo* microscopy for each subset of subjects stratified by colposcopic impression and histological diagnosis. Of the 59 women in the study who had no colposcopic lesions, only three (5%) showed high grade disease (CIN 2+) by histopathology. HRME correctly classified all three (100%) of those patients as abnormal, and correctly classified 45 of the remaining 56 (80%) as normal. In addition, among the 85 women who had lesions noted by colposcopy, 47 (55%) showed high grade disease (CIN 2+) by histopathology. HRME imaging correctly classified 45 (96%) of this subset of patients as abnormal, and correctly classified 16 of the remaining 38 (42%) in the subset as normal.

Overall, HRME identified as positive 20 out of 21 (95%) of patients with CIN 2, 24 out of 25 (96%) of patients with CIN 3, and 4 out of 4 (100%) of patients with invasive cancer; and identified as negative 27 out of 43 (63%) of patients with CIN 1 and 34 out of 51 (67%) of patients with benign histopathology (**Table 3-4**). Colposcopy identified as

positive 20 out of 21 (95%) of patients with CIN 2, 23 out of 25 (92%) of patients with CIN 3, and 4 out of 4 (100%) of patients with invasive cancer; and identified as negative 20 out of 43 (47%) of patients with CIN 1 and 36 out of 51 (71%) of patients with benign histopathology (**Table 3-4**).

		Per-site Analysis (n = 229)		Per-patient Analysis (n = 144)	
Colposcopic Impression	Histological Diagnosis	No. of sites measured	No. (%) of sites HRME positive	No. of patients measured	No. (%) of patients HRME positive
Normal	Benign	51	24 (47%)	36	9 (25%)
	CIN 1	21	3 (14%)	20	2 (10%)
	CIN 2	2	2 (100%)	1	1 (100%)
	CIN 3	3	3 (100%)	2	2 (100%)
Abnormal	Benign	36	24 (67%)	15	8 (53%)
	CIN1	36	24 (67%)	23	14 (61%)
	CIN 2	37	33 (89%)	20	19 (95%)
	CIN 3	36	33 (92%)	23	22 (96%)
	AIS/Invasive Carcinoma	7	7 (100%)	4	4 (100%)

Table 3-4 *In vivo* microscopy results stratified by colposcopic impression and histological diagnosis on a per-site and per-patient basis.

Clinically normal sites where no biopsy was taken were excluded from the analysis because those sites lack a histopathologic gold standard. Only sites with a histopathologic gold standard were included in order to make the analysis as rigorous as possible. If clinically normal sites with no biopsy are included in the analysis, colposcopic examination must be used as the gold standard for those sites. This may be done, and it results in improved specificity for HRME imaging (HRME: 92% sensitivity, 81% specificity on a per-site basis).

3.5 Discussion

The primary findings of our study include a 37% relative and 23% absolute increase in the diagnostic completion follow-up rate of screen-positive women

attending a mobile diagnostic van equipped with colposcopy and *in vivo* microscopy in their local community compared with those asked to travel to a central hospital for diagnosis. The diagnostic completion rate in the central hospital arm (considered baseline) was 64%, and the diagnostic completion rate in the mobile van arm was 87%. The absolute increase represents the difference between those percentages (87%-64% = 23%) and the relative increase represents the percentage increase relative to baseline $(87\%-64\%)/64\% = 37\%$, using unrounded values in the calculation). Furthermore, we found the combination of colposcopy and point-of-care *in vivo* microscopy to be feasible in a mobile van.

Current approaches for cervical cancer screening in high-income countries include Pap and / or HPV DNA testing.⁴⁶ Patients with abnormal results undergo colposcopy with cervical biopsies and if clinically significant precursor lesions are identified, ablative (cryotherapy) or excisional procedures such as CKC or LEEP are performed. Although these screening and diagnosis algorithms are very effective, they are expensive and require high-level infrastructure and well-trained personnel. In addition, they require two or more separate patient visits with communication of test results between visits. These strategies therefore often fail in lower-resource settings such as underserved areas of Brazil and other LMICs where there is often a lack of trained personnel, infrastructure and pathology services.⁷ Furthermore, geographical distances and cultural barriers result in many women with abnormal screening tests not receiving the recommended diagnostic and treatment procedures because they are unable to travel to central healthcare facilities for the multiple necessary follow-up visits.³⁰ As a result, many women present with invasive disease often at an advanced stage when curative treatment is no longer achievable.

In the current study, HRME was found to have similar performance characteristics to colposcopy when using histopathologic diagnosis as the gold standard. On a per-patient basis, HRME had a sensitivity of 96% and specificity of 65% while colposcopy had a sensitivity of 94% and specificity of 60%. The analyses by cervical lesion site showed similar findings. In addition, HRME correctly identified 95% of the CIN 2 lesions, 96% of the CIN 3 lesions and 100% of the invasive cancer lesions. These results are similar to previous study findings from this group evaluating the HRME device.^{11,12} Pierce *et al.*¹¹ reported on 174 women in rural China who underwent cervical cancer screening with HPV testing, visual inspection with acetic acid (VIA), colposcopy and HRME imaging. Of the 69 women noted to have abnormalities on colposcopy, only 12 (17%) showed high-grade disease on biopsy. However, HRME imaging correctly classified all 12 high-grade areas (100%) as abnormal, and correctly classified 38 of the remaining 57 (67%) as normal. Furthermore, when patients were stratified based on a positive high-risk HPV DNA test, HRME imaging correctly identified 100% of the patients with CIN 2+.¹¹ A subsequent report by Grant *et al.*¹² evaluated 59 women in Brazil undergoing colposcopy for abnormal screening Pap test results. In addition to colposcopy, all participants also underwent HRME imaging. Biopsies were obtained of abnormal areas and the histopathologic results were compared to the colposcopy and HRME findings. HRME was found to have a sensitivity of 92% and specificity of 77% compared with histopathologic diagnosis. The HRME device has also been shown to be effective as a diagnostic tool in both oral cancer and esophageal cancer screening.^{13,36-38}

This study was appropriately powered to detect a 30% difference in diagnostic completion rate. Because actual cluster size was smaller than the cluster size used in

sample size calculations, the *post-hoc* statistical power was >95% over a very wide range of intracluster correlation coefficients.

When using the HRME, as with any point probe device, the clinician must choose where to place the probe. Colposcopy is one method to guide probe placement, but visual inspection can also be used. In this study the HRME probe was placed (a) on colposcopically abnormal sites (guided by colposcopy), and (b) on at least one site in each quadrant even if no colposcopic lesion was present (guided by visual examination).

The potential role of the HRME in the clinical setting depends on the other diagnostic resources that are available. In settings where colposcopy and histopathology are routinely available, the use of HRME in combination with colposcopy could potentially help guide the selection of biopsy sites, reduce the number of biopsies needed for diagnosis, and enable see-and-treat strategies. In settings where colposcopy and histopathology are not routinely available (and treatment decisions are often made on the basis of methods known to have poor specificity such as VIA), the use of HRME could potentially help improve specificity and reduce overtreatment.

One limitation of the HRME in this study was that images of acceptable quality were unable to be acquired for a small number of patients and cervical sites. The automated QC check failed in 6% of cervical sites imaged. Notably, over half of these QC failures occurred in a single very hot day early in the study in which the air conditioning unit within the mobile diagnostic van was not functioning, causing the HRME camera to overheat over the course of the day (a problem subsequently avoided simply by ensuring that the device was not left on continuously throughout the day). Another limitation of HRME is that it is not designed to sample the endocervix. Of the

45 patients that were both colposcopy negative and HRME negative, there were no CIN 2+ cervical biopsies. However, 2 of these 45 women did have a positive ECC result (CIN 3 in both cases), and were referred for treatment accordingly. Finally, HRME imaging requires the use of the fluorescent contrast agent proflavine. There were no adverse events related to proflavine exposure in this study; long term follow-up studies of participants exposed to proflavine are ongoing.

One limitation of the study design is that the cost effectiveness and patient satisfaction of the mobile diagnostic van were not assessed. Additionally, all diagnostic and treatment procedures were performed by physicians fully trained as expert colposcopists, which would not be feasible in this setting outside of a research study.

In summary, our findings show that women referred to the mobile van equipped with colposcopy and *in vivo* microscopy were statistically significantly more likely to complete their diagnostic follow-up than women referred to a central hospital, without any compromise in clinical care. *In vivo* microscopy provides an automated, point-of-care diagnostic imaging capability that can readily be implemented in a mobile van, with sensitivity and specificity similar to colposcopy.

Our results provide important information regarding the feasibility of this approach. Larger prospective studies evaluating HPV primary screening followed by HRME are ongoing in Brazil, El Salvador and in the Rio Grande Valley along the Texas-Mexico border. These studies will address some of the above limitations including the feasibility of performing HRME imaging in a variety of clinical settings and by non-specialists including primary care physicians, nurse practitioners, midwives and physician assistants.

4. IS PROFLAVINE EXPOSURE ASSOCIATED WITH DISEASE PROGRESSION IN WOMEN WITH CERVICAL DYSPLASIA? A BRIEF REPORT

The contents of this chapter have been published in the following journal article: Pantano N, Hunt B, Schwarz RA, Parra S, Cherry K, Possati-Resende JC, Longatto-Filho A, Fregnani JH, Castle PE, Schmeler K, Richards-Kortum R. Is proflavine exposure associated with disease progression in women with cervical dysplasia? A brief report. *Photochemistry and photobiology* 94.6 (2018): 1308-1313.

4.1 Abstract

Proflavine is an acridine dye used with high-resolution microendoscopy for *in vivo* diagnostic evaluation of cervical epithelial cells. However, there are concerns that even short-term exposure of cervical tissue to dilute proflavine may increase cervical cancer risk. We performed a retrospective analysis of women referred for colposcopy to Barretos Cancer Hospital comparing the risk of cervical disease progression in those whose cervical tissue was (n=232) or was not exposed (n=160) to proflavine. Patients in both groups underwent treatment and follow-up based on histopathologic results and per the local standards of care. Progression of disease was evaluated by comparing histopathology from the initial visit to the worst subsequent histopathology result from all follow-up visits. Mean duration of follow-up was 18.7 and 20.1 months for the proflavine-exposed and controls groups, respectively. There were no significant differences in disease progression from normal/CIN 1 to CIN 2/3 or from any initial diagnosis to invasive cancer between the proflavine-exposed and control groups overall. Risks of cervical dysplasia progression observed in this study are in agreement with those of the natural history of cervical cancer. Our results suggest that cervical exposure to dilute proflavine does not increase the risk of cervical precancer and cancer.

4.2 Introduction

Proflavine is a fluorescent dye that has long been recognized and used as a topical antibacterial agent. Its history of use as an antiseptic in wound care dates back to the early 1900s, and its action as a nucleic acid intercalator has been established as the basis of its antibacterial properties^{47,48}. Today proflavine continues to be used as a commercially available antiseptic in many parts of the world^{49,50}. Proflavine is one of the components of triple dye, which is routinely used for umbilical cord care in newborn infants in the United States^{51,52}. Toxicity of triple dye is rare⁵².

With the development of *in vivo* fiber optic microscopy techniques over the past two decades, proflavine and acriflavine (another closely related acridine compound) have been frequently used as topical contrast agents for optical imaging due to their ability to fluorescently label cell nuclei with high contrast. Acriflavine has been reported as a contrast agent for *in vivo* confocal laser endomicroscopy in the colon^{53,54}, stomach⁵⁵, duodenum⁵⁶, upper gastrointestinal tract⁵⁷, and central airway⁵⁸. In these studies, acriflavine was applied topically to the tissue prior to imaging, usually in combination with intravenous fluorescein. Our group has reported the use of topically applied proflavine as a contrast agent for *in vivo* microscopy in the oral cavity³⁶, esophagus³⁷, and cervix¹². The use of proflavine or acriflavine in combination with *in vivo* microscopy enables real-time assessment of the morphology and distribution of cell nuclei, aiding in the accurate identification of pre-cancerous lesions⁵⁹.

Despite their long history of safe clinical use, there remains disagreement on the cancer risk of using proflavine and acriflavine as contrast agents for *in vivo* imaging. In 1977, the National Cancer Institute published a bioassay of proflavine for possible carcinogenicity, based on administration of proflavine in the diet to groups of rats and mice over a two-year period⁶⁰. The bioassay was inconclusive due to an unusually high

incidence of carcinomas in control animals. The International Agency for Research on Cancer (IARC) published evaluations of the carcinogenic risk of acriflavine in 1977⁶¹ and of proflavine in 1980⁶². These evaluations concluded that while acriflavine and proflavine display mutagenicity *in vitro*, their carcinogenic risk to humans could not be classified due to inadequate data in experimental animals and a lack of data in humans. Some view the mutagenicity of these compounds *in vitro* as sufficient reason for concern. For example, a recent publication recommends the use of methylene blue over acriflavine as an imaging contrast agent, due to concerns about the *in vitro* mutagenicity of acriflavine⁶³. However, other recent publications suggest that acriflavine and proflavine exhibit beneficial anti-cancer and anti-viral effects. Acriflavine has been shown to reduce tumor growth in mice by inhibiting HIF-1 dimerization⁶⁴. An acriflavine and proflavine mixture has been shown to elicit an anti-viral immune response that significantly reduced rhinovirus infection in mammalian cells⁶⁵. At this time, there is no evidence demonstrating the carcinogenicity of proflavine or acriflavine in human subjects, and as such, both compounds remain categorized by the IARC as “not classifiable as to its carcinogenicity to humans”^{66,67}.

Yet concerns remain. Therefore, we conducted a retrospective study to evaluate whether proflavine exposure is associated with progression of cervical neoplasia in women with abnormal cervical cytology by comparing the rate of cervical intraepithelial neoplasia progression between two groups, one exposed to proflavine and the other not exposed.

4.3 Materials and methods

4.3.1 Study participants

This is a historical cohort study. The proflavine group was composed of women who had previously participated in a clinical trial (ID# NCT02335372) at Barretos Cancer Hospital (Brazil) between June 2013 and January 2015. Participants in the trial were recruited from women presenting for colposcopy due to abnormal cytology or previous history of cervical dysplasia at the Prevention Department of Barretos Cancer Hospital. This trial evaluated a new optical imaging device, the high resolution microendoscope (HRME), a portable, battery-powered fluorescence microscope with a flexible fiber-optic probe ^{13,34,68}. The details of this experimental device have previously been described ¹². As part of the study protocol, 5% acetic acid was applied to the surface of the uterine cervix and standard of care colposcopy was performed. This was followed by application of 5% Lugol's iodine solution. Proflavine solution at a concentration of 0.01% was then applied to the cervix followed by evaluation with the HRME probe. The typical volume dispensed by the spray bottle was measured to be 3 mL with a range of 1 to 7 mL. The control group was identified retrospectively and was balanced for severity of baseline cytology diagnosis prior to colposcopy. The control group comprised of women who underwent colposcopy and cervical biopsies in the Prevention Department of Barretos Cancer hospital between May 2013 and May 2016. Patients in the control group did not undergo evaluation with HRME and thus were not exposed to proflavine; however, they were subjected to equivalent diagnostic, treatment, and follow-up procedures. This retrospective study was approved by the Barretos Cancer Hospital Ethics Research Committee, the Brazilian National Ethics Research Commission (CAAE: 83227718.8.0000.5437).

4.3.2 Follow-up and treatment

Follow-up data including all histopathology results were collected for patients in both groups from the time of the initial colposcopy until the last appointment available from the medical record. In both groups, the treatment and follow-up were performed according to the local standard of care based on histopathology results. Those women diagnosed with cervical intraepithelial neoplasia grades 2 (CIN 2) or 3 (CIN 3) underwent treatment with loop electrosurgical excision procedure (LEEP). Those diagnosed with invasive cancer were referred to the Gynecologic Oncology department for care.

4.3.3 Statistical analysis

Data for all patients in the proflavine and control groups were collected from electronic medical records at Barretos Cancer Hospital and compiled into an SPSS database file. The database consisted of a total 499 patient records with 299 records in the proflavine group and 200 in the control group. The following data fields were aggregated for each participant: age, study group, initial cytology result, date of initial colposcopy, initial histopathology result, date of last clinical visit, and up to six subsequent histopathology results. For all histopathology results the date of diagnosis and tissue specimen type (cervical biopsy, endocervical curettage (ECC), LEEP, or hysterectomy) were documented. The total follow-up time period was calculated as the time between the first and last colposcopy evaluations on record.

Cervical cytology results were grouped into two categories: normal/low-grade and high-grade. Normal/low-grade included the following: negative for intraepithelial lesions or malignancy (NILM), atypical squamous cells of undetermined significance (ASC-US), and low-grade squamous intraepithelial lesion (LSIL). High-grade cytology results included: high-grade squamous intraepithelial lesion (HSIL), atypical squamous

cells cannot exclude high-grade squamous intraepithelial lesion (ASC-H), atypical glandular cells (AGC), squamous cell carcinoma, and adenocarcinoma. In order to assess cervical dysplasia progression, histopathology results were grouped into five clinically relevant categories by increasing severity of diagnosis: negative for intraepithelial lesion (NIL), cervical intraepithelial neoplasia grade 1 (CIN 1), CIN 2 or CIN 3 (CIN 2/3), AIS, and invasive cancer. NIL category included diagnoses of normal, inflammation, hyperplasia, and metaplasia. CIN 2/3 category included diagnoses of CIN 2, CIN 3, and CIN 2/3. Invasive cancer category included diagnoses of invasive squamous cell carcinoma and adenocarcinoma. In cases where follow-up colposcopy was negative and no biopsies were obtained, the visit was categorized as normal colposcopy/no biopsy performed.

A total of 107 patients were excluded from the analysis for the following reasons: 1) total follow-up duration was less than 6 months (proflavine group, n = 49; control group, n = 29), and 2) patient had an initial histological diagnosis of invasive carcinoma (proflavine group, n = 18; control group, n = 11). After applying these exclusion criteria, 392 patients were included in the retrospective analysis (proflavine group, n = 232; control group, n = 160). Using these data the study population was characterized using descriptive statistics. Categorical variables were compared using Fisher's exact test and continuous numerical variables were compared using a Student's t-test. Survival analysis was performed on two clinical endpoints: 1) progression from a baseline histopathologic diagnosis of <CIN 2 to a subsequent histopathologic diagnosis of CIN 2+, 2) treatment via LEEP. A spreadsheet containing all patients was compiled with the following information: cohort (proflavine or control), time from baseline diagnosis to the event of interest (days), and the status (1=event occurred, 0=event not occurred) at the indicated time interval. A survival function was fit using the 'survival' package of

the R programming language⁶⁹. The functions for both proflavine and control cohorts were plotted over time and the resulting fits compared using the G-rho family of tests⁷⁰. Statistical significance level was set at 5%. Post-hoc statistical power was calculated using PASS (NCSS LLC, Kaysville, Utah, USA).

4.4 Results

4.4.1 Baseline patient characteristics

There were no statistically significant differences between groups according to their initial diagnosis by cytology and / or histopathology from cervical biopsies (**Table 4-1**). The mean age at diagnosis was slightly higher in the control arm (mean=37.8 years; sd=13.0 years) in comparison to the proflavine arm (mean=35.8 years; sd=11.7 years) but not statistically significant (p=0.10). The mean time of follow-up was 18.7 months (sd=6.0 months) and 20.1 months (sd=5.6 months) for the proflavine and control groups, respectively.

	Study group		p-value [†]
	Proflavine-exposed (n=232)	Control (n=160)	
Cytology			
Normal/low-grade	54 (23.3%)	36 (22.5%)	0.90
High-grade	178 (76.7%)	124 (77.5%)	
Histopathology			
< CIN2	125 (53.9%)	77 (48.1%)	0.30
CIN2+	107 (46.1%)	83 (51.9%)	
Normal/low-grade	: NILM, ASC-US, LSIL		
High-grade	: ASC-H, AGC, HSIL, Carcinoma, Adenocarcinoma		
< CIN2	: Cervicitis, Hyperplasia, Metaplasia, CIN1		
CIN2+	: CIN 2, CIN 3, Adenocarcinoma In-situ, Carcinoma, Adenocarcinoma		

[†] p-values calculated using Fisher's exact test

Table 4-1 Baseline diagnosis according to the study group.

4.4.2 Cervical dysplasia progression

In order to evaluate cervical dysplasia progression, we compared initial and subsequent worst histopathologic diagnoses for the patients exposed to proflavine ($n=232$) and those not exposed ($n=160$). Because removal of proflavine-exposed tissue via LEEP may mitigate the effect of proflavine exposure, we stratified our analysis within each cohort by those who underwent LEEP and those who did not. A LEEP was performed in 124 of the 232 (53%) patients exposed to proflavine and in 96 of the 160 (60%) control patients ($p=0.21$).

		Exposed to proflavine and did not undergo LEEP ($n = 108/232$, 47%)						Not exposed to proflavine and did not undergo LEEP ($n = 64/160$, 40%)						
		Worst Subsequent Pathologic Diagnosis						Worst Subsequent Pathologic Diagnosis						
Initial Pathologic Diagnosis	Normal Colposcopy/No Biopsy	NIL	CIN 1	CIN 2/3	AIS	Invasive Cancer	Total	Normal Colposcopy/No Biopsy	NIL	CIN 1	CIN 2/3	AIS	Invasive Cancer	Total
	NIL	10 (9%)	2 (2%)	5 (5%)	0 (0%)	0 (0%)	17 (16%)	17 (27%)	7 (11%)	4 (6%)	0 (0%)	0 (0%)	0 (0%)	28 (44%)
	CIN 1	45 (42%)	12 (11%)	28 (26%)	0 (0%)	0 (0%)	85 (79%)	19 (30%)	2 (3%)	7 (11%)	1 (2%)	0 (0%)	0 (0%)	29 (45%)
	CIN 2/3	1 (1%)	0 (0%)	2 (2%)	3 (3%)	0 (0%)	0 (0%)	6 (6%)	1 (2%)	1 (2%)	1 (2%)	4 (6%)	0 (0%)	7 (11%)
	AIS	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
	Total	56 (52%)	14 (13%)	35 (32%)	3 (3%)	0 (0%)	0 (0%)	108 (100%)	37 (58%)	10 (16%)	12 (19%)	5 (8%)	0 (0%)	64 (100%)

 Cases where subsequent pathology was equal to initial

 Cases where subsequent pathology was worse than initial

- Normal Colposcopy/No Biopsy : Follow-up colposcopy examinations were normal and did not require additional biopsies
- NIL : Negative for Intraepithelial Lesion (normal, inflammation, hyperplasia, and metaplasia)
- CIN 1 : Cervical Intraepithelial Neoplasia Grade 1
- CIN 2/3 : Cervical Intraepithelial Neoplasia Grades 2, 3, and 2/3
- AIS : Adenocarcinoma In-situ
- Invasive Cancer : Invasive adenocarcinoma

Table 4-2 Initial and worst subsequent pathologic diagnoses for patients who did not undergo LEEP.

Table 4-2 shows the initial and subsequent worst diagnoses for patients who did not undergo LEEP in the proflavine and control groups respectively. There were a total of nine patients that had benign histopathology who had a subsequent diagnosis of CIN 1 (proflavine 5%; control 6%; $p=0.73$). Additionally, there was one patient in the control group that progressed from an initial diagnosis of CIN 1 to CIN 2/3 (proflavine 0%;

control 2%). The majority of non-LEEP patients had no abnormal cervical lesions observed during follow-up colposcopy and thus no additional cervical biopsies were required (proflavine 52%; control 58%; p=0.53).

		Exposed to proflavine and underwent LEEP (n = 124/232, 53%)						Not exposed to proflavine and underwent LEEP (n = 96/160, 60%)							
		Worst Subsequent Pathologic Diagnosis						Worst Subsequent Pathologic Diagnosis							
Initial Pathologic Diagnosis	Normal Colposcopy/No Biopsy	NIL	CIN 1	CIN 2/3	AIS	Invasive Cancer	Total	Normal Colposcopy/No Biopsy	NIL	CIN 1	CIN 2/3	AIS	Invasive Cancer	Total	
	NIL	0 (0%)	1 (1%)	0 (0%)	6 (5%)	0 (0%)	0 (0%)	7 (6%)	0 (0%)	0 (0%)	3 (3%)	4 (4%)	0 (0%)	1 (1%)	8 (8%)
	CIN 1	0 (0%)	0 (0%)	7 (6%)	9 (7%)	0 (0%)	0 (0%)	16 (13%)	0 (0%)	0 (0%)	3 (3%)	9 (9%)	0 (0%)	0 (0%)	12 (13%)
	CIN 2/3	0 (0%)	8 (6%)	11 (9%)	81 (65%)	0 (0%)	0 (0%)	100 (81%)	0 (0%)	5 (5%)	8 (8%)	63 (66%)	0 (0%)	0 (0%)	76 (79%)
	AIS	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (1%)	1 (1%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
	Total	0 (0%)	9 (7%)	18 (15%)	96 (77%)	0 (0%)	1 (1%)	124 (100%)	0 (0%)	5 (5%)	14 (15%)	76 (79%)	0 (0%)	1 (1%)	96 (100%)

 Cases where subsequent pathology was equal to initial

 Cases where subsequent pathology was worse than initial

- Normal Colposcopy/No Biopsy : Follow-up colposcopy examinations were normal and did not require additional biopsies
- NIL : Negative for Intraepithelial Lesion (normal, inflammation, hyperplasia, and metaplasia)
- CIN 1 : Cervical Intraepithelial Neoplasia Grade 1
- CIN 2/3 : Cervical Intraepithelial Neoplasia Grades 2, 3, and 2/3
- AIS : Adenocarcinoma In-situ
- Invasive Cancer : Invasive adenocarcinoma

Table 4-3 Initial and worst subsequent pathologic diagnoses for patients who underwent LEEP.

Table 4-3 provides a similar summary of initial and final histopathology for patients in both study groups that did undergo LEEP during the follow-up period. The most common initial biopsy diagnosis in both patient populations was CIN 2/3, which was subsequently equally confirmed by the diagnosis on the LEEP tissue (proflavine 65%; control 66%; p=1.0). There were 28 patients with benign initial histopathology on biopsy who had a worst subsequent diagnosis of CIN 2/3 on the LEEP tissue (proflavine 12%; control 13%). In addition, three patients in the control group had benign initial histopathology with a worst subsequent diagnosis of CIN 1 (proflavine 0%; control 3%). One invasive cervical cancer was diagnosed both in the proflavine-exposed and control groups.

The risks of cervical dysplasia progression for all groups are summarized in

Table 4-4. Progression from normal/CIN 1 to CIN 2/3 was noted in 15/232 (6.5%) proflavine-exposed patients and 14/160 (8.8%) of controls ($p=0.44$). Progression from any initial diagnosis to invasive cancer was noted in one of 232 patients exposed to proflavine (0.4%) and one of 160 controls (0.6%) ($p=1.0$). Based on the total number of cases where subsequent pathology was worse than initial, there were no statistically significant differences in progression of cervical dysplasia between proflavine-exposed and non-exposed patients (**Table 4-4**) (all patients: $p=0.19$; non-LEEP patients: $p=0.50$; LEEP patients: $p=0.35$).

Initial Diagnosis	Worst Subsequent Diagnosis	All Patients		p-value [†]	Did Not Undergo LEEP		p-value	Underwent LEEP		p-value
		Proflavine exposure (n = 232)	No exposure (n = 160)		Proflavine exposure (n = 108)	No exposure (n = 64)		Proflavine exposure (n = 124)	No exposure (n = 96)	
NIL	CIN 1	5 (2.2%)	7 (4.4%)	0.19	5 (4.6%)	4 (6.3%)	0.50	0 (0%)	3 (3.1%)	0.35
NIL	CIN 2/3	6 (2.6%)	4 (2.5%)		0 (0%)	0 (0%)		6 (4.8%)	4 (4.2%)	
CIN 1	CIN 2/3	9 (3.9%)	10 (6.3%)		0 (0%)	1 (1.6%)		9 (7.3%)	9 (9.4%)	
NIL	Invasive Cancer	0 (0%)	1 (0.6%)		0 (0%)	0 (0%)		0 (0%)	1 (1.0%)	
AIS	Invasive Cancer	1 (0.4%)	0 (0%)		0 (0%)	0 (0%)		1 (0.8%)	0 (0%)	
Total		21 (9.1%)	22 (13.8%)		5 (4.6%)	5 (7.8%)		16 (12.9%)	17 (17.7%)	

[†] p-values calculated using Fisher's exact test based on total cases where subsequent diagnosis was worse than initial

Table 4-4 Summary of cases where subsequent pathologic diagnosis was worse than baseline diagnosis

Figure 4-1 shows a Kaplan-Meier estimate of two clinical endpoints: progression from a baseline histopathologic diagnosis of <CIN 2 to a subsequent histopathologic diagnosis of CIN 2+ (proflavine group, $n = 125$; control group, $n = 77$) and treatment via LEEP (proflavine group, $n = 232$; control group, $n = 160$). No statistically significant difference was found for progression of histological diagnosis from <CIN 2 to CIN 2+ between proflavine-exposed and control groups (log-rank test, $p=0.21$). Additionally, no

statistically significant difference was found for treatment by LEEP between proflavine-exposed and control groups (log-rank test, $p=0.32$).

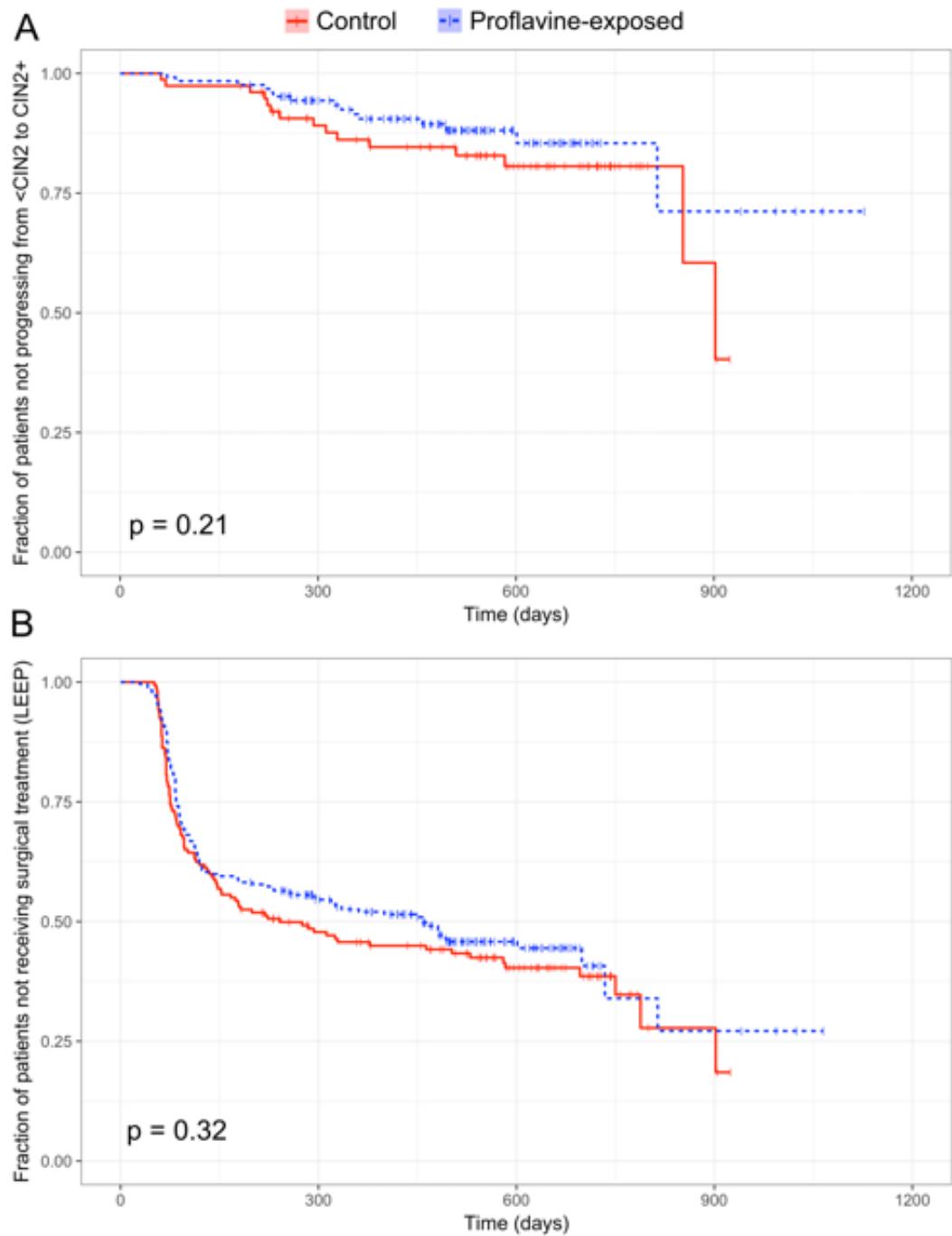


Figure 4-1 Kaplan-Meier estimate of A) progression from <CIN 2 to CIN 2+ and B) surgical treatment (LEEP). Vertical tick marks represent censoring at the indicated time point. Statistical comparisons were performed using a log-rank test.

4.5 Discussion

In this study we assessed whether proflavine might increase the risk of cervical neoplasia progression when used as a fluorescent contrast agent for *in vivo* microscopy during colposcopic examination. Our results showed no significant increases in cervical dysplasia progression in women exposed to proflavine when compared with a control group of non-exposed women. This is important for routine gynecological exams because proflavine can be used as a diagnostic tool in conjunction with colposcopy, facilitating the recognition of cervical lesions through *in vivo* microscopy as previously reported by Grant *et al*¹². The primary contribution of this study is to provide initial data regarding the use of proflavine in human subjects with cervical dysplasia. We are not aware of any prior studies assessing proflavine and cervical dysplasia progression; however, these results may also be put into context of other studies investigating the natural history of cervical dysplasia.

In a meta-analysis of 17 studies involving 4,504 participants, Östör found that 10% of women with CIN 1 will progress to CIN 3 and only 1% will progress to invasive cancer⁷¹. Follow-up durations in this meta-analysis varied widely from as little as three weeks up to 25 years⁷¹. A more recent analysis of 524 patients from the placebo arm of the quadrivalent HPV vaccine trials found that 12% of patients with initial diagnosis of CIN 1 developed CIN 2/3 during follow-up⁷². The findings in this study are in agreement with previously established risks.

The total follow-up duration achieved by this analysis is one limitation of the study. Even with this limitation, the structure of this study provides a meaningful comparison for cervical dysplasia progression in proflavine-exposed and non-exposed subjects. The post-hoc statistical power was ≥80% (alpha = 0.05) to detect an absolute

increase in risk of upstaging from <CIN 2 to CIN 2+ of 18% (or more). However, longer-term follow-up studies with larger sample sizes will be necessary to provide reassurance for the topical use of proflavine on the cervical epithelium.

In conclusion, the risks of cervical dysplasia progression observed in this study are in agreement with those of the natural history of cervical cancer. Our findings suggest that cervical proflavine exposure is not associated with acute disease progression in women with cervical dysplasia.

5. CERVICAL LESION ASSESSMENT USING REAL-TIME MICROENDOSCOPY IMAGE ANALYSIS: THE CLARA STUDY

5.1 Abstract

Background: High-resolution microendoscopy (HRME) is a promising method for real-time diagnosis of pre-cancerous cervical lesions at the point-of-care. The objective of the study was to prospectively evaluate diagnostic performance of HRME in women with abnormal cervical cancer screening tests.

Methods: Women referred for colposcopy at Barretos Cancer Hospital were offered enrollment into the study. All participants underwent colposcopy, HRME, and cervical biopsies with histopathologic diagnosis. Using consensus histopathology diagnosis as the gold standard, the sensitivity and specificity of colposcopy and HRME were compared using endpoints of cervical intraepithelial neoplasia grade 2 or more severe (CIN 2+) and grade 3 or more severe (CIN 3+).

Findings: Of 1,486 subjects included in the study analysis, there were 435 (29%) CIN 2+ cases and 341 (23%) CIN 3+ cases. The sensitivity to detect CIN 2+ was 95.6% (95%CI: 93.2–97.2%) for colposcopy vs. 91.7% (95%CI: 88.7–94.0%) for HRME ($p < 0.01$); with a specificity of 63.4% (95%CI: 60.4–66.2%) for colposcopy vs. 59.7% (95%CI: 56.7–62.6%) for HRME ($p = 0.02$). The sensitivity to detect CIN 3+ was 96.2% (95%CI: 93.5–97.8%) for colposcopy vs. 95.6% (95%CI: 92.8–97.4%) for HRME ($p = 0.81$); with a specificity of 58.7% (95%CI: 55.8–61.5%) for colposcopy vs. 56.6% (95%CI: 53.7–59.4%) for HRME ($p = 0.18$).

Interpretation: HRME was shown to have comparable sensitivity and specificity to colposcopy guided cervical biopsy for the detection of CIN 3+. Our results suggest that

HRME could provide a low-cost, point-of-care alternative to colposcopy and biopsy, allowing for cervical screening, diagnosis and treatment of in a single visit.

5.2 Introduction

Cervical cancer remains a major global health concern with low-resource and rural areas being disproportionately affected.⁷³ Routine procedures used in high-resource settings such as cytology, colposcopy, and cervical biopsy with histopathology are costly and difficult to implement and scale in low-resource settings.^{7,74} In addition, the multiple visits required lead to a high loss to follow-up for women with abnormal screening results.⁷⁵ In 2018, the World Health Organization (WHO) called for the elimination cervical cancer as a public health problem. The proposed strategy includes that 70% of women are screened with a HPV test at 35 and 45 years of age.⁷⁴ Primary screening using HPV testing has great promise to aid in identifying women at risk for cervical cancer but most women who test HPV positive do not have high-grade cervical abnormalities or cancer. Therefore, effective strategies for triage and appropriate treatment of screen-positive women in low-resource settings are urgently needed.^{76,77}

In Brazil and many other countries, the standard of care for screen-positive women is to undergo colposcopic evaluation of the cervix using magnifying optics by a trained medical provider. If any areas of the cervix appear visually abnormal during colposcopy, the provider performs a biopsy of the tissue and it is submitted for histopathologic examination by another highly trained provider. However, availability of colposcopy and pathology services are often limited in low-resource settings, resulting in long delays between visits and loss to follow-up for women in need of preventive treatment.⁷⁴

Cervical cancer is a leading cause of death in underserved populations of Brazil.⁷⁸

Barretos Cancer Hospital (BCH) is a rural Brazilian hospital tasked with providing cancer prevention to a large, rural geographic region. They have implemented innovative mobile screening programs to improve access to screening^{27,30,32}, but diagnostic follow-up of screen-positive women is challenging. In a previous study at BCH, we observed that over one-third of women with abnormal screening test did not attend their follow-up visit when being referred for colposcopy.⁷⁹

In vivo microscopy has the potential to transform clinical decision-making by performing microscopic examination of biological tissues in real time, obviating the need for colposcopy and cervical biopsy.⁸⁰ We developed a low-cost, high-resolution microendoscope (HRME) capable of imaging cervical epithelium with sub-cellular resolution.³⁴ Previous smaller clinical studies have shown HRME to be feasible and accurate for detection of high-grade cervical abnormalities in China, Botswana, Brazil, Texas, and El Salvador.^{10-12,79,81} The objective of this study was to conduct a large scale, prospective assessment of the diagnostic performance of HRME in a high-risk population.

5.3 Materials and methods

5.3.1 Study design and participants

We conducted a non-blinded, single arm trial to prospectively assess the diagnostic performance of HRME to detect high-grade cervical abnormalities in a high-risk, rural population with abnormal screening tests.

Potential subjects were identified through a regional screening program operated by BCH. Per Brazilian guidelines, women 18 years of age or older who had an abnormal screening test (\geq ASC-US cytology: atypical squamous cells of undetermined

significance or more abnormal; hrHPV+: positive for any high-risk HPV types included in the cobas 4800 HPV test) were referred for colposcopy. Women scheduled for colposcopy at BCH were assessed for eligibility using the following inclusion criteria: 1) positive cervical screening test (cytology \geq ASC-US or hrHPV+), 2) at least 18 years old, 3) intact cervix, 4) not pregnant (negative urine pregnancy test required for subjects of child-bearing potential), 5) no known allergy to proflavine or acriflavine, 6) do not belong to an indigenous Brazilian population, 7) willing and able to provide written informed consent. Women who met these inclusion criteria were invited to participate in the study.

This study was approved by the BCH Ethics Research Committee, the Brazilian National Ethics Research Commission (CAAE: 61743416.1.0000.5437), and the Institutional Review Boards of Rice University (ID#2017-293) and The University of Texas MD Anderson Cancer Center (ID#2017-0096). Written informed consent was obtained from all participants. The protocol was registered with ClinicalTrials.gov (NCT03195218).

5.3.2 High-resolution microendoscope

The HRME is a portable, fiber-optic fluorescence microscope designed to obtain high-resolution images of cervical tissue *in vivo* with sub-cellular resolution.^{13,34,82} The system consists of the following components: the microscope hardware, flexible imaging probe, foot-pedal to trigger image acquisition, and graphical user interface deployed on a PC tablet.¹³ Images with a 790 micron field-of-view and 4 micron lateral spatial resolution are obtained by placing the fiber optic probe (Fujikura FIGH-30-850N , Myriad Fiber Imaging Tech. Inc., NJ, USA) in gentle contact with the cervix.

The HRME is used following topical application of the fluorescent antiseptic proflavine (0.01% concentration). Proflavine stains cell nuclei in the superficial epithelium, enabling real-time assessment of the size, shape, and distribution of cell nuclei. Proflavine has a long history of safe clinical use.^{47,50,52} A retrospective case-control analysis of screen-positive women previously exposed to proflavine during HRME imaging found no significant differences in disease progression between the proflavine exposed and control groups.⁸³

Per BCH regulations, imaging probes were disinfected prior to each use as follows: water rinse, 15 minutes in enzymatic detergent (Aniosyme DLT Plus, Laboratoires Anios, France), water rinse, 30 minutes in Cidex OPA (Cilag GmbH International, Switzerland), water rinse, and wipe with 70% alcohol. Probes were stored in a sealed container until next use. Regular maintenance of HRME instruments (i.e. imaging of calibration targets to ensure optical alignment and replacement of optical probes as needed) was performed every 6–12 months throughout the study.

The HRME software allows clinicians to acquire and automatically analyze images to determine whether the imaged region is likely to contain high-grade cervical abnormalities. The diagnostic algorithm was developed previously in a pilot study at BCH, and uses morphological analysis to segment and categorize individual cell nuclei within the image as normal or abnormal.⁷⁹ The software calculates the number of abnormal nuclei per square millimeter, which is used to categorize the overall image as non-neoplastic or neoplastic based on a pre-established threshold value of 120 abnormal nuclei/mm². The hardware, software, algorithm, and algorithm thresholds were fixed at the beginning of the study.

5.3.3 Diagnostic exam procedure

All diagnostic exams and study procedures were performed by one of three study colposcopists in the Cancer Prevention department at BCH. Exam rooms were equipped with a table for gynecologic examination, a colposcope, and an HRME system (**Figure 5-1**).

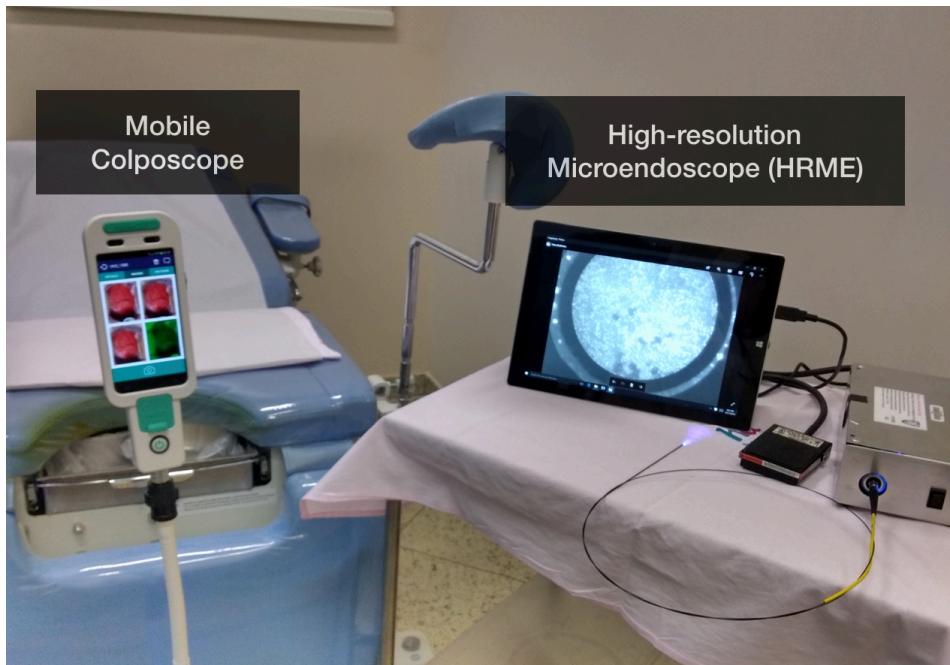


Figure 5-1 Photograph of high-resolution microendoscope and mobile colposcope setup in clinic.

The clinician first performed a general visual inspection of the cervix with a cervical swab sample collection. Colposcopy was then performed with application of 5% acetic acid followed by Lugol's iodine. Colposcopy images were acquired both after application of each of these solutions. Clock positions of any colposcopically abnormal areas were indicated on the study case report form (CRF) along with the colposcopist's impression (low-grade, high-grade, or suspected cancer).

Following colposcopy, proflavine at a concentration of 0.01% was applied to the cervix and four-quadrant HRME imaging was performed. Within each quadrant, HRME images were acquired from each area that was noted to be abnormal by

colposcopy. If no lesions were noted by colposcopy, HRME images were acquired from a colposcopically normal site at the squamocolumnar junction. For each site imaged with the HRME, the following information was recorded on the CRF: site number, clock position, HRME result (number of abnormal nuclei per square millimeter and diagnosis), colposcopist impression of tissue type (squamous, columnar, or metaplasia), and colposcopy impression (normal, low-grade, high-grade, or suspected cancer).

Upon completion of HRME imaging, sites identified as abnormal by colposcopy and/or by HRME imaging were biopsied. In cases where a large lesion spanning multiple quadrants was present, colposcopists biopsied the most severe lesion site based on colposcopy impression. If no abnormal sites were identified by either method, then a single biopsy was taken from a normal site randomly selected by the clinician. Endocervical curettage (ECC) was performed if indicated per local standard of care.

5.3.4 Histopathology & follow-up

Histopathology was used as the gold standard to assess the diagnostic performance of colposcopy and HRME imaging for detection of high-grade cervical abnormalities. For each study case, two experienced pathologists independently reviewed the histologic slides from each biopsy. They categorized their overall diagnosis for each biopsy specimen into one of twelve categories established at the outset of the study. The classifications follow the standard criteria outlined by the World Health Organization³⁹ and included the following: atrophy, inflammation, metaplasia, cervical intraepithelial neoplasia (CIN) 1, CIN 2, CIN 3, AIS, squamous cell carcinoma, adenocarcinoma, adenosquamous carcinoma, indeterminate (insufficient for diagnosis), or 'other' in which case the pathologist specified a classification category not included in this list. Discrepant results were resolved by consensus review between the

two pathologists. Based on final histopathology results, participants were then treated or scheduled for follow-up per standard of care. This included excision with loop electrosurgical excision (LEEP) or cold knife conization (CKC) for women with CIN 2/3 and referral to gynecologic oncology for those diagnosed with AIS or invasive cancer.

5.3.5 Statistical analyses

The primary endpoint was a comparison of the diagnostic performance of HRME and colposcopy for detection of high-grade cervical abnormalities. Sample size calculations were performed beforehand to determine the number of subjects needed to estimate 95% confidence intervals of both sensitivity and specificity within a 4% margin of error. Assumptions regarding prevalence of high-grade cervical abnormalities and sensitivity / specificity of HRME were based on a previous pilot study using HRME at BCH.⁷⁹ Assuming a CIN 2+ prevalence of 25%, a 12.5% drop-out rate, and a 4% margin of error, the target accrual was determined to be 1,600 subjects.

Study data were first collected on a paper case report form and subsequently entered into a study database with corresponding electronic forms using REDCap (Research Electronic Data Capture).⁴² De-identified colposcopy and HRME images were uploaded to the study database for review and verification of all study cases. Additionally, de-identified lab report documents were uploaded for all other diagnostic test results including cytology, HPV testing, and histopathology.

Pathology diagnoses were grouped into five categories: negative (including acanthosis, atrophy, inflammation, metaplasia), CIN 3, CIN 2, CIN 3 (including CIN 3 and AIS), and invasive carcinoma (including squamous cell carcinoma, adenocarcinoma, adenosquamous carcinoma, and carcinosarcoma). Both CIN 2 or more severe diagnoses (CIN 2+) and CIN 3 or more severe diagnoses (CIN 3+) were

used as endpoints to evaluate the diagnostic performance of colposcopy and HRME. HRME was recorded as a binary result (positive or negative) and colposcopy positivity was defined as low-grade or more severe impression per the standard of care.

Summary reports including all patient outcomes, per-site diagnostic results, and pathology findings were exported directly from REDCap. These data were directly imported into Mathematica 12.0 for cross-tabulation of contingency tables and calculation of descriptive statistics (frequency counts, mean, and range). Contingency tables were then entered into GraphPad Prism 8.3 for further statistical analyses and generation of graphical visualizations.

Sensitivity and specificity with 95% confidence intervals (95%CI) were calculated on both a per-site basis and a per-patient basis, with the most abnormal colposcopy result, HRME result, and histopathology result used in the per-patient analysis. 95%CI of sensitivity and specificity were generated the modified Wald method.⁸⁴ Significance testing comparing the sensitivity and specificity of colposcopy and HRME was performed using McNemar's test.⁴⁵ Two additional analyses of the HRME quantitative diagnostic parameter (number abnormal nuclei per mm²) were performed. First, one-way analysis of variance (ANOVA) on ranks was performed with Dunn's multiple comparison tests to evaluate differences in HRME scores with increasing severity of pathology diagnosis. Second, receiver operator characteristic (ROC) analysis was performed using all possible thresholds between individual HRME scores to calculate sensitivity /specificity achieved when the threshold for positivity by HRME is varied. Area under the ROC curve was calculated using the trapezoidal rule.

5.4 Results

Figure 5-2A shows the number of participants at each stage of the study. A total of 2,028 women with abnormal screening tests were interviewed to assess eligibility prior to their scheduled colposcopy. Of those, 1,821 (89.9% of 2,028) were deemed eligible and invited to participate in the study. Two hundred twenty-one subjects declined to participate, and the remaining 1,600 (87.9% of 1,821) provided written informed consent. A total of 1,523 subjects completed diagnostic examination including HRME imaging. Thirty-seven participants were excluded from the analysis for the following reasons: 35 with indeterminate histopathology results, one patient revoked consent, and one patient completed the exam but had missing HRME images. The remaining 1,486 (92.9%) participants with complete diagnostic information were included in the analysis.

Figure 5-2B summarizes the age, cytology and HPV results, and final histopathology diagnosis for subjects included in the analysis. The mean age of participants was 40.0 years old (SD=12.1), with 769 (54%) between the ages of 30 to 49. The distribution of pathology diagnoses was as follows: 882 (59%) negative, 169 (11%) CIN 1, 94 (6%) CIN 2, 319 (21%) CIN 3, and 22 (1%) cancer. The prevalence of high-grade cervical abnormalities (CIN 2+) in the analysis cohort was 29%.

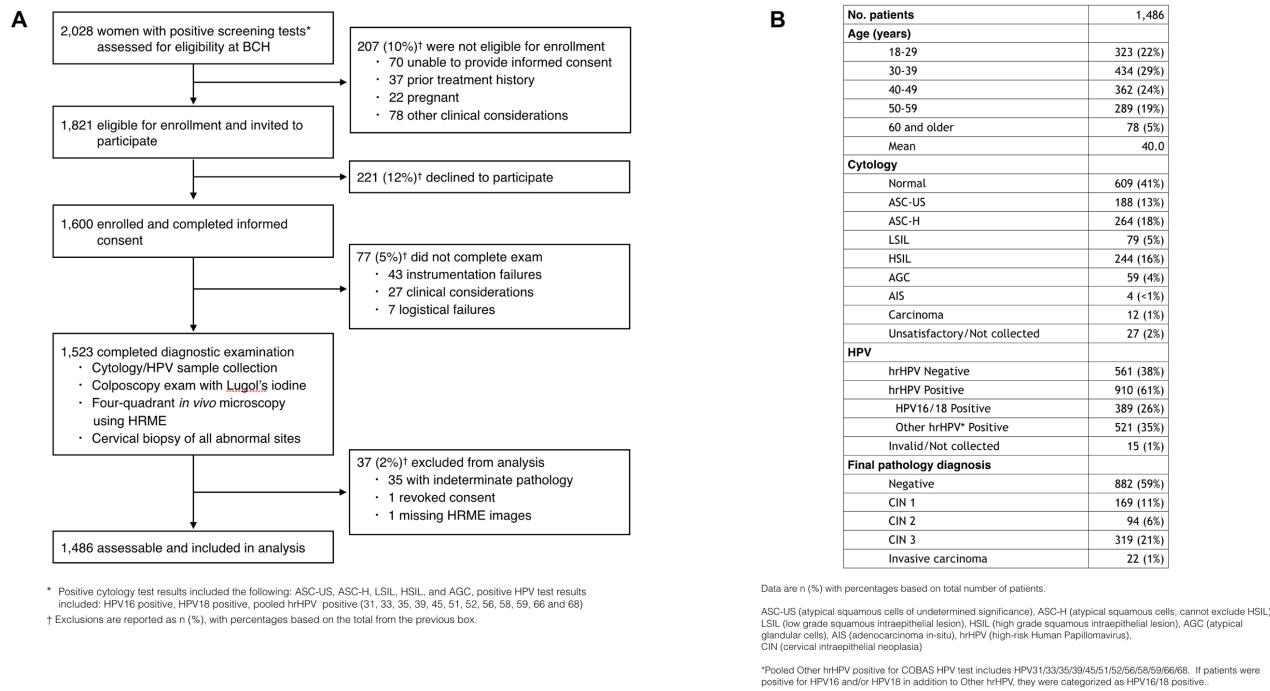


Figure 5-2: Per patient analysis dataset. **A)** Flow diagram of eligibility screening, enrollment, and analysis subsets of study participants. **B)** Age, Cytology / HPV testing results, and final pathology diagnosis for 1,486 patients included in analysis.

Figure 5-3 shows the diagnostic performance of colposcopy and four-quadrant HRME using histopathology as the gold standard. Diagnostic performance was assessed using either CIN 2+ or CIN 3+ as threshold for positivity. **Figure 5-3A** provides a contingency table showing the agreement of colposcopy (low-grade or more severe) and HRME by histopathology result for all participants. Percent agreement between colposcopy and HRME was 79%. **Figure 5-3B** is a scatter plot of HRME scores by histopathology result. Mean HRME scores increased with increasing severity of cervical abnormalities (one-way ANOVA on ranks with Dunn's multiple comparison tests: Neg vs CIN 1: p<0.001, CIN 1 vs CIN 2: p=0.002, CIN 2 vs CIN 3: p<0.001, CIN 3 vs Inv Ca: p=0.41).

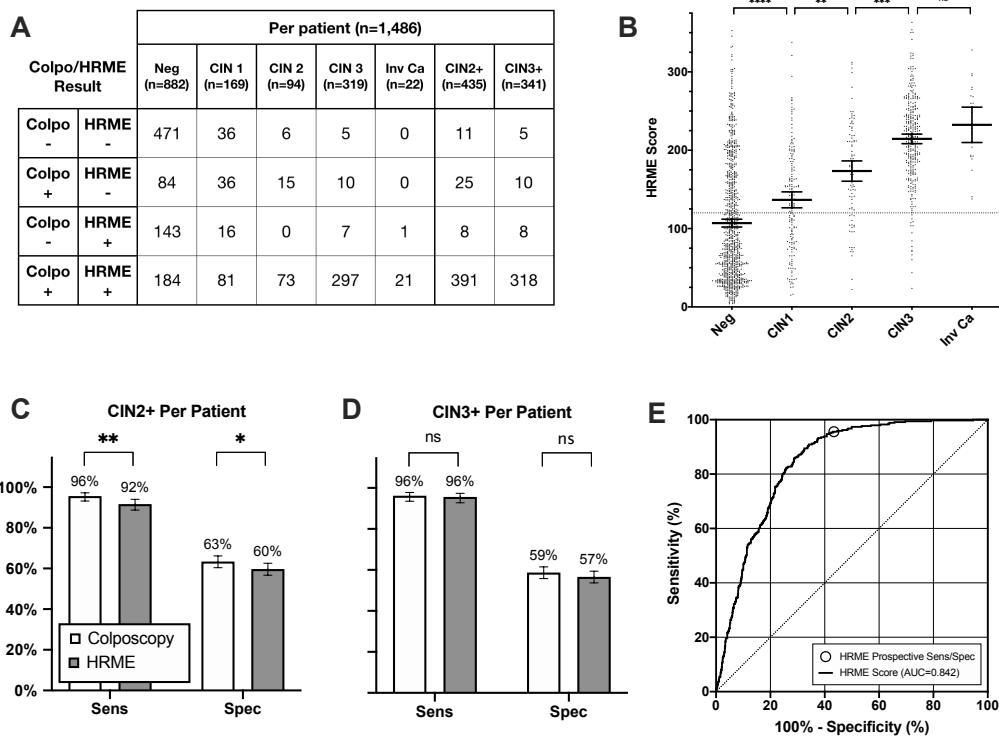


Figure 5-3: Primary study outcome. Diagnostic performance of colposcopy and HRME using histopathology as the gold standard. **A)** Table showing the correlation of colposcopy (low-grade or more severe) and HRME by pathology result for each patient. **B)** Dot scatter plot of per-patient HRME scores, the quantitative HRME diagnostic parameter (number of abnormal nuclei per square millimeter), stratified by pathology. The dotted line represents the prospective cutoff for positivity in the study. Bars overlaid on scatter plots indicate mean and 95% confidence intervals within each column. Statistical comparisons between column means was performed using Dunn's test. **C-D)** Comparisons of the sensitivity (Sens) and specificity (Spec) of colposcopy and four-quadrant HRME for detection of both CIN 2+ and CIN 3+. Error bars represent 95% confidence intervals. Significance testing was performed using McNemar's test. **E)** ROC curve of HRME score for detection of CIN 3+.

Figures 5-3C and 5-3D contain bar charts comparing the sensitivity and specificity of colposcopy and HRME for histologic endpoints of CIN 2+ and CIN 3+. HRME had comparable sensitivity and specificity to colposcopy for detection of CIN 3+, while having slightly lower sensitivity and specificity for detection of CIN 2+. The sensitivity of colposcopy for CIN 2+ detection was 95.6% (95%CI: 93.2–97.2%) vs .91.7% (95%CI: 88.7–94.0%) for HRME ($p < 0.01$, McNemar's test). The specificity for CIN 2+ detection for colposcopy was 63.4% (95%CI: 60.4–66.2%) vs. 59.7% (95%CI: 56.7–62.6%)

for HRME ($p = 0.02$). The sensitivity for CIN 3+ detection for colposcopy was 96.2% (95%CI: 93.5–97.8%) vs. 95.6% (95%CI: 92.8–97.4%) for HRME ($p = 0.81$). The specificity for CIN 3+ detection for colposcopy was 58.7% (95%CI: 55.8–61.5%) vs. 56.6% (95%CI: 53.7–59.4%) for HRME ($p = 0.18$). **Figure 5-3E** is a receiver operator characteristic (ROC) plot showing the tradeoff between sensitivity and specificity as the diagnostic threshold of HRME positivity is varied; the area under the curve for detection of CIN 3+ was 0.842.

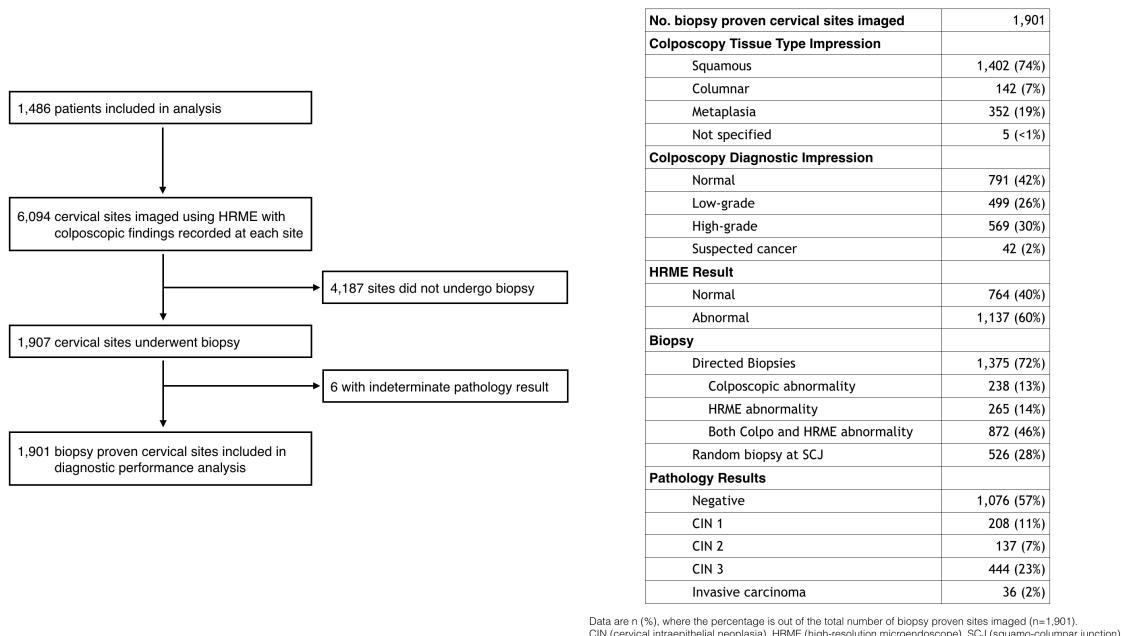


Figure 5-4: Per site analysis dataset. **A)** Flow diagram of cervical sites imaged and included in per site analysis. **B)** Colposcopy, HRME, and pathology results for 1,901 biopsy proven cervical sites.

Because some participants underwent multiple biopsies, a per biopsy-site analysis in which diagnostic performance was evaluated on individual biopsy results was also conducted (**Figure 5-4A**). The per site distribution of colposcopy impressions, HRME results, and histopathology results are provided (**Figure 5-4B**). Again, HRME had comparable sensitivity and specificity to colposcopy for detection of CIN 3+, while having slightly lower sensitivity and specificity for detection of CIN 2+ (**Figure 5-5**).

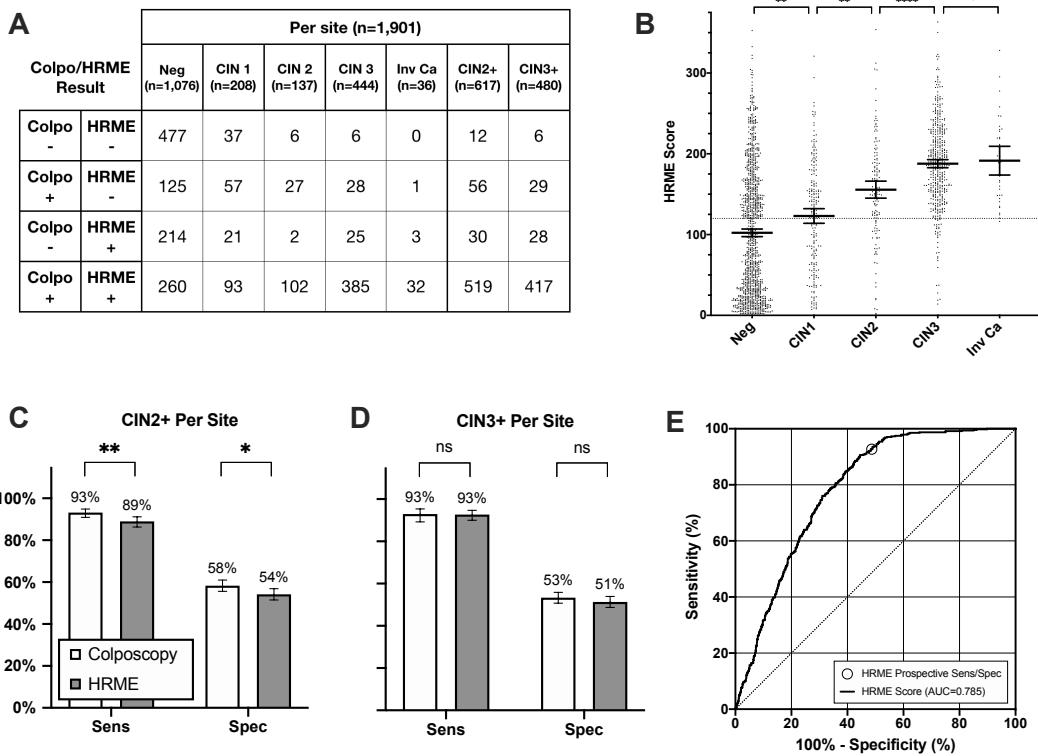


Figure 5-5: Primary study outcome using a per biopsy site analysis instead of a per patient analysis. Diagnostic performance of HRME for 1,901 biopsy proven cervical sites imaged. **A)** Table showing the correlation of colposcopy and HRME by pathology result for biopsied site. **B)** Dot scatter plot of per-patient HRME scores stratified by pathology. The dotted line represents the prospective cutoff for positivity in the study. Bars overlaid on scatter plots indicate mean and 95% confidence intervals within each column. Statistical comparisons between column means was performed using Dunn's test. **C-D)** Comparisons of the sensitivity (Sens) and specificity (Spec) of colposcopy and HRME for detection of both CIN 2+ and CIN 3+. Error bars represent 95% confidence intervals. Significance testing was performed using McNemar's test. **E)** ROC curve of HRME score for detection of CIN 3+.

Differences in diagnostic performance of HRME by tissue type (squamous, columnar, and metaplasia) were assessed using tissue type impressions provided by colposcopy. Among 1,901 biopsied sites the distribution of tissue type impressions were as follows: 1,402 (74%) squamous, 142 (7%) columnar, 352 (19%) metaplasia, 5 (<1%) not specified (**Figure 5-4B**). A stratified analysis of HRME performance was performed for the 1,402 squamous sites compared to 494 non-squamous (columnar / metaplasia) sites (**Figure 5-6**). HRME scores were significantly higher among those with a non-squamous impression for sites with negative and CIN 1 biopsy results (negative: Dunn's test,

$p<0.001$; CIN 1: Dunn's test, $p<0.05$) (**Figure 5-6A**). The elevated scores observed in non-squamous sites resulted in decreased ability to discriminate sites with high-grade abnormalities (squamous vs non-squamous: CIN 2+, 0.80 vs 0.62 AUC; CIN3 +, 0.81 vs 0.67 AUC) (**Figure 5-6B-C**).

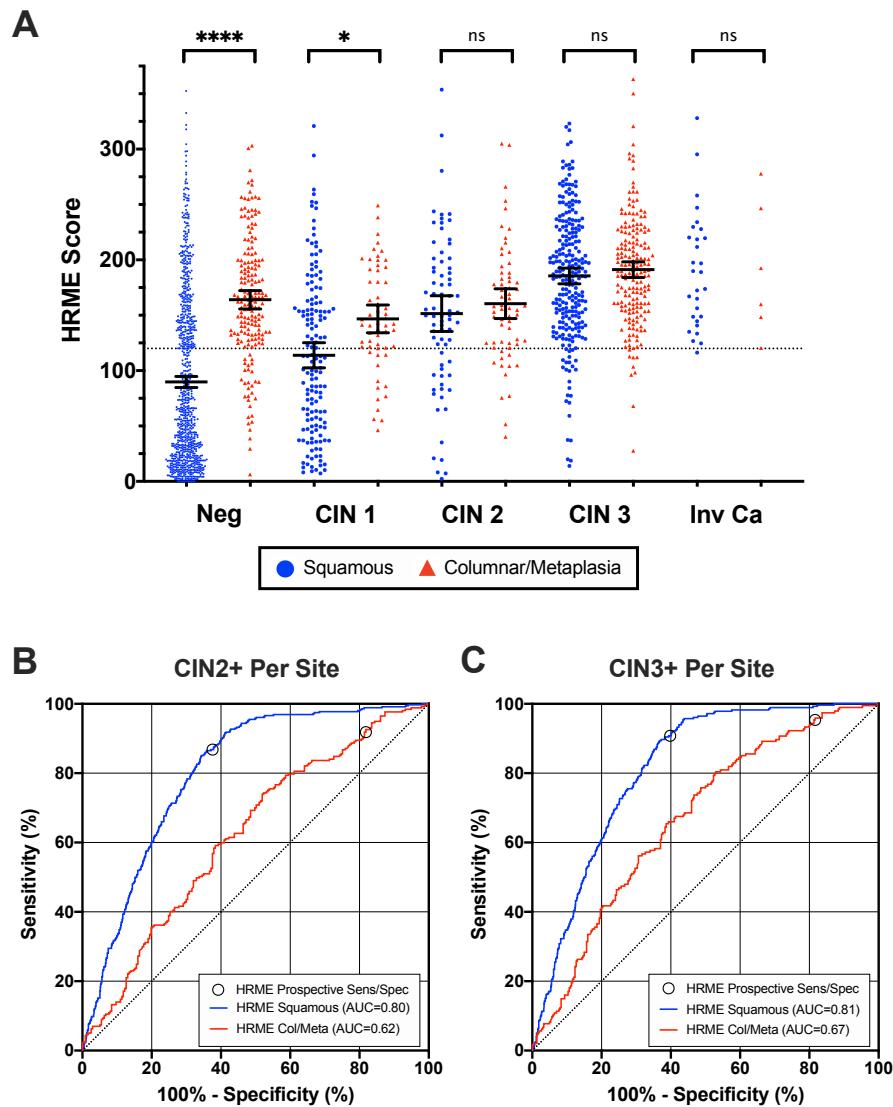


Figure 5-6: Comparison of HRME performance between sites with squamous and non-squamous (columnar/metaplasia) tissue type impressions. **A)** Dot scatter plot of per-site HRME scores stratified by pathology and tissue type impression. The dotted line represents the prospective cutoff for positivity in the study. Bars overlaid on scatter plots indicate mean and 95% confidence intervals within each column. Statistical comparisons between column means was performed using Dunn's test. **B-C)** ROC curves of HRME score for detection of CIN 2+ and CIN 3+ stratified by squamous and non-squamous impressions.

Figure 5-7 presents colposcopy and HRME images from five patients cases including: 1) a CIN 2 case in which HRME and colposcopy were both positive; 2) a CIN 2 case in which HRME was falsely negative; 3) a CIN 3 case in which colposcopy was falsely negative, 4) a negative case in which both HRME and colposcopy were falsely positive, and 5) a carcinoma case in which HRME and colposcopy were both positive. In each case, a cervical diagram summarizes the colposcopy impressions, HRME sites and results, and histopathology results.

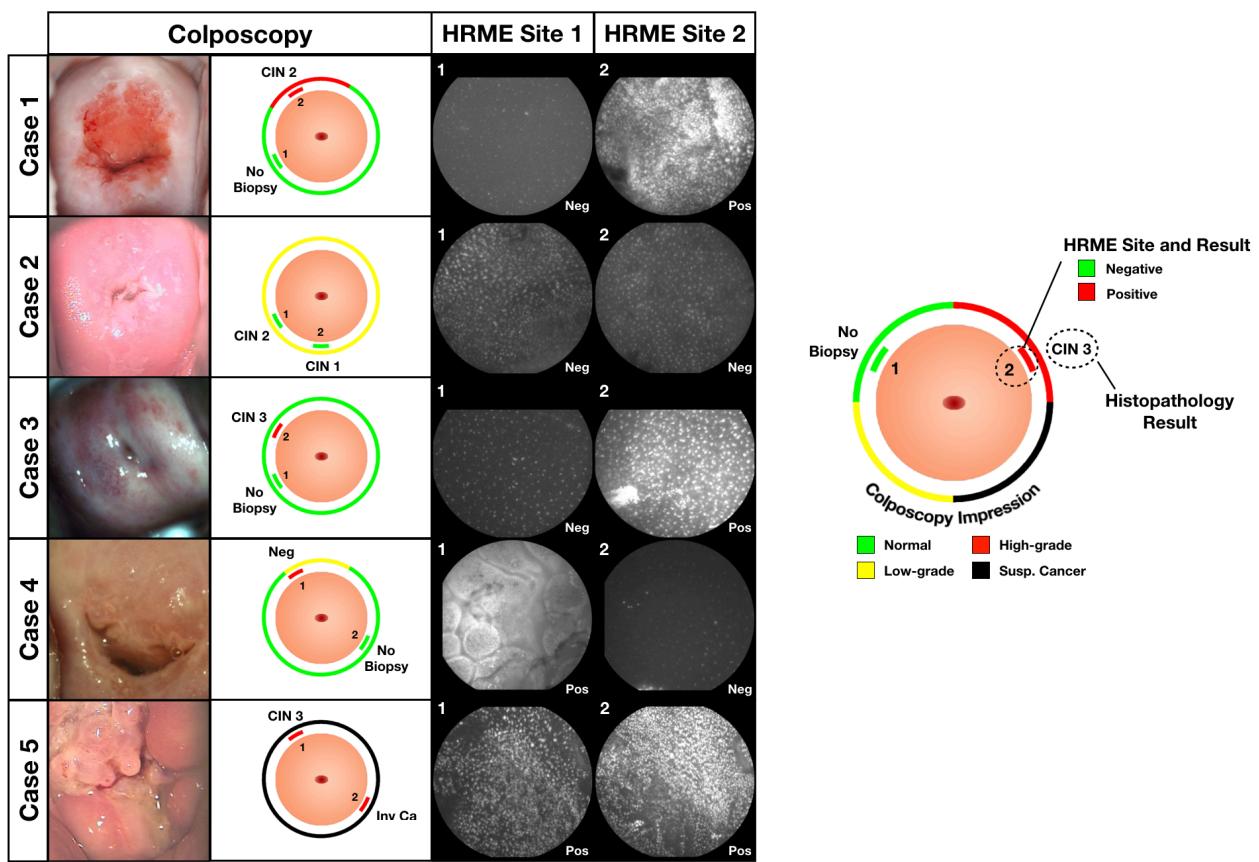


Figure 5-7: Case example with images and diagnostic outcomes for five subjects.

In the first case, the colposcopist noted a high-grade lesion from 11 to 1 o'clock. HRME imaging was negative in three of the four quadrants, but positive at 11 o'clock. The HRME image at the positive site contained enlarged, pleomorphic, and densely

crowded nuclei. A single punch biopsy was taken at 11 o'clock which was diagnosed as CIN 2 by histopathology.

In the second case, the colposcopist noted a 360° low-grade lesion. HRME images acquired at 2 o'clock, 5 o'clock, 6 o'clock, 9 o'clock, and 11 o'clock were all negative. The HRME image at 8 o'clock demonstrated some nuclei crowding and pleomorphism which resulted in an elevated HRME score that was still below the diagnostic threshold (104 out of 120). Two punch biopsies were taken at 6 o'clock and 8 o'clock which were diagnosed as CIN 1 and CIN 2 respectively.

In the third case, no lesions were observed during colposcopy. During four-quadrant HRME imaging, the 10 o'clock site demonstrated enlarged nuclei with moderate crowding and pleomorphism. This was the only site positive by HRME or colposcopy. A single punch biopsy was taken at 10 o'clock resulting in a diagnosis of CIN 3.

In fourth case, the colposcopist noted a low-grade lesion from 11 to 12 o'clock near the cervical os. Squamous epithelium was imaged in three of the four quadrants and the resulting HRME images displayed characteristics of normal squamous epithelium. The fourth HRME image was acquired near the cervical os where the colposcopist noted a low-grade lesion and appeared visually distinct from the other sites. The tissue type impression of the colposcopist at this site was columnar epithelium. The image analysis algorithm failed to accurately segment the dense nuclei structures in the image, and resulted in a positive result by HRME. Histopathology diagnosis of the biopsy taken at 11 o'clock was negative for cervical abnormalities and characterized as chronic cervicitis with immature squamous metaplasia of the endocervical epithelium.

In the final case, an abnormal growth was observed throughout the cervix and the colposcopist suspected the patient had developed cancer. Two punch biopsies were taken at 5 and 11 o'clock. HRME images from the biopsied sites contained enlarged, pleomorphic nuclei and were abnormal by the real-time image analysis algorithm. Histopathology analysis of the 5 o'clock biopsy confirmed the presence of invasive squamous cell carcinoma.

5.5 Discussion

The primary finding of our study is that automated diagnosis using HRME image analysis performed comparably to the standard of care (colposcopy guided biopsy) for detection of CIN 3+. We observed a small but statistically significant difference between colposcopy and HRME for detection of CIN 2+. At BCH, CIN 2+ is the threshold for deciding which patients will undergo preventative treatment; however, CIN 2 is often considered equivocal disease and treatment of CIN 2 may not be the ideal strategy all settings. We observed a strong trend of increasing HRME score with increasing pathology diagnosis severity (Neg, CIN 1, CIN 2, CIN 3, Invasive Carcinoma). Further optimization of automated HRME diagnostic algorithms for detection of CIN 3+ maybe even more suitable for single visit approaches to avoid over-treatment of patients with CIN 2. An important distinction between HRME and the standard of care is that HRME provides a result at the point-of-care, and thus could potentially reduce loss to follow-up.

Pathology-based diagnosis is a significant challenge to many low- and middle-income countries (LMICs) in terms of infrastructure and human resources. In southeastern Brazil, there are an estimated 968 pathologists for 87,500,000 people (one per 90,000 people).⁸⁵ In sub-Saharan Africa, there is less than one pathologist per 500,000

people.⁸⁶ A recent review by Wu and colleagues highlighted pathology services as a barrier to addressing cervical cancer in LMICs, noting a multitude of factors that make it difficult to implement including: "training requirements for clinical pathologists and laboratory technicians, sufficient infrastructure, equipment, maintenance contracts, and reagents to properly transport, fix, and process tissue for histologic analysis".⁸⁷ Real-time microendoscopy image analysis is a new technology that can overcome many of the infrastructural and human resource constraints of conventional pathology services. Previous studies at BCH and in the Texas-Mexico border have demonstrated the ability for HRME to be successfully utilized in mobile clinics and by non-specialist healthcare providers.^{79,81}

While demonstrating great potential for providing real-time diagnosis, there are some important limitations to HRME and this study. First, HRME is unable to sample the endocervical canal. In 30 study cases, the colposcopist suspected an endocervical lesion and performed an endocervical curettage. Of those, 23 were negative and 7 were positive (23% CIN 2+). In 2 out of the 7 positive cases endocervical curettage identified a high-grade abnormality that was not identified by HRME imaging of the ectocervix. Second, in order to compare performance of colposcopy and HRME a per-biopsy-site basis, HRME probe placement within each quadrant of the cervix was guided to lesions identified by colposcopy. Lastly, punch biopsies specimens analyzed by histopathology were larger than the field of view of individual HRME images, so there is not a perfect correlation between the tissue reviewed for pathology diagnosis and the tissue imaged with the HRME. Future studies could acquire colposcopy video of HRME probe placement to more rigorously document image and biopsy locations for correlation with histopathology findings.

Some areas for further development and improvement of HRME

instrumentation were identified during the course of this validation study. First, the lifetime of optical probes should be extended. Optical probes were replaced every 6-12 months during the study due to breakage. Repeated exposure to disinfection detergents required per BCH guidelines is hypothesized to have played a role degrading the outer coating of the optical probes and leading to increased brittleness. Strategies to increase durability and total lifetime of the optical probes, such as a protective outer casing, can help reduce per test costs associated with HRME. Second, sustained exposure to elevated humidity contributed to degradation of one of the optical filters within the HRME. Filter degradation reduced image quality in a small number of patient cases before the problem was identified. Monitoring filter quality with replacement at regular intervals was sufficient to solve the problem, but a better long-term solution is needed. Recent work by Long and colleagues noted the need for more low-cost and rugged optical filtering components, highlighting a novel material that could be used in the future.⁸⁸ Lastly, the stratified analysis of HRME performance demonstrated that benign columnar and metaplasia epithelium are challenging to distinguish using current HRME image analysis approaches. However, this study has generated a significant library of biopsy proven HRME and colposcopy images. These data can enable deep learning techniques to be utilized and further improve automated diagnostic strategies for HRME. Colposcopy images acquired can be used to validate automated visual evaluation algorithms, which are a promising alternative to colposcopy guided HRME probe placement.¹⁸ Strategies that combine automated image analysis of HRME and colposcopy images are a promising future research direction.

Accurate, real-time diagnostics that can aid in management of screen positive women are greatly needed. This will become even more urgent, particularly in rural

healthcare settings, as governments and healthcare systems strive to meet the WHO target of screening 70% of women twice in their lifetime. Our study highlights the potential to use *in vivo* microscopy for real-time diagnosis to guide immediate treatment and overcome losses to follow-up in low-resource settings.

6. BENCHMARKING PERFORMANCE OF CONVOLUTIONAL NEURAL NETWORKS FOR DIAGNOSTIC CLASSIFICATION OF CERVICAL DYSPLASIA USING THE CLARA STUDY DATASET

The contents of this chapter will be included as part of publication to be submitted.

6.1 Overview

Automated image interpretation of high-resolution microendoscopy images can provide real-time decision support for single visit approaches to cervical cancer prevention. Current strategies for HRME image analysis use traditional computer vision (TCV) techniques to segment and quantify shape and size of cell nuclei within images. While this approach has demonstrated good prospective performance for detection of high-grade cervical abnormalities, the image features and parameters utilized by the algorithm include many manually supervised parameters and present limited opportunity for model optimization. Developing generalizable image classifiers using this approach becomes increasingly challenging with large datasets. Recently, convolutional neural networks have demonstrated improved performance for several medical image classification tasks. A key aspect of the strategy is end-to-end optimization of model parameters through gradient-based learning of labeled examples. In this study, I utilized the datasets from the clinical studies in Chapters 3 and 5 of this thesis to evaluate the capability of existing convolutional neural networks to classify HRME images of cervical tissue.

6.2 Materials and methods

The purpose of this study was to evaluate the performance of an existing convolutional neural network to analyze high-resolution microendoscopy images of cervical tissue and detect the presence of high-grade cervical abnormalities. This

approach is compared to a previously developed image analysis method that uses traditional computer vision techniques (i.e. morphological operations) to segment and classify nuclei as ‘normal’ and ‘abnormal’ based on their area and eccentricity (**Figure 6.1**). The previously established approach is referred to as the traditional computer vision (TCV) algorithm.

6.2.1 Image dataset

The data utilized in the present study comes from two clinical trials conducted at Barretos Cancer Hospital (BCH) in Barretos, Brazil. These studies were approved by the Brazilian National Ethics Research Commission (CAAE #s: 61743416.1.0000.5437 and 37774314.3.0000.5437) and the institutional review boards of Rice University (ID #s: 2015–0442 and 2017–293) and the University of Texas MD Anderson Cancer Center (ID#s: 2015–0442 and 2017–0096). Written informed consent was obtained from participants.

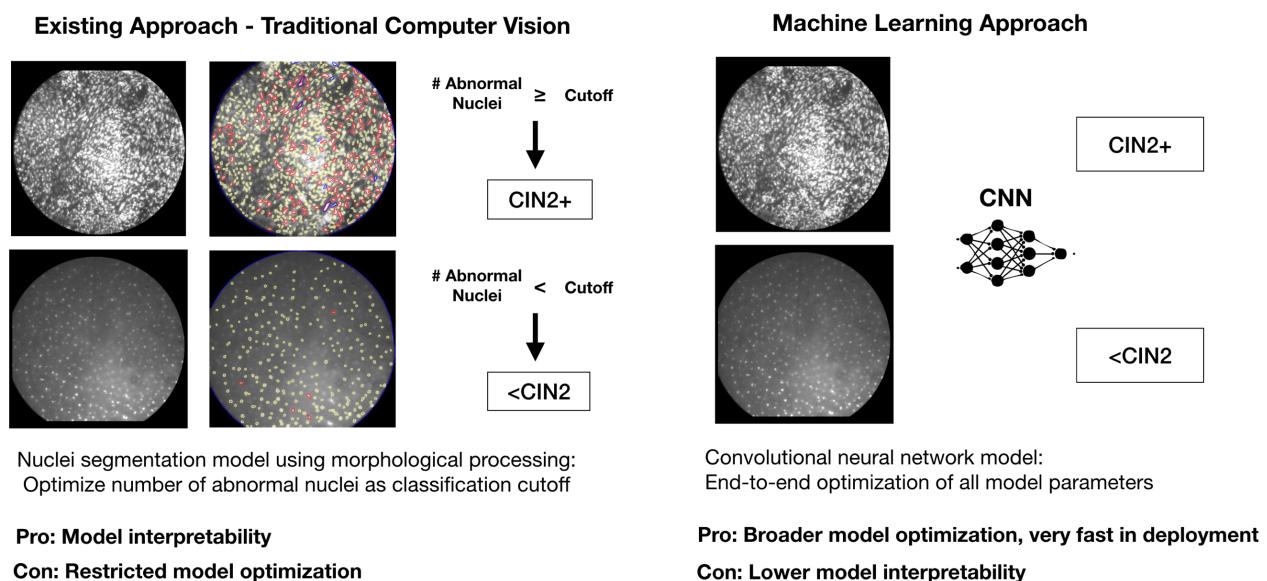


Figure 6-1 Comparison of the existing approach using traditional computer vision to a machine learning approach for diagnostic prediction of cervical intraepithelial grade two or more severe (CIN 2+).

A total of N = 1,627 participants from both studies underwent colposcopy, four-quadrant high-resolution microendoscopy with real-time image analysis using the existing traditional computer vision algorithm, and tissue biopsies with histopathology diagnosis. 1,535 out of 1,627 (94%) patients underwent diagnostic examination in the Cancer Prevention Department at BCH. The remaining 92 (6%) underwent diagnostic examination in a mobile diagnostic van in their community. All patients were seen by one of three study colposcopists that participated in both studies. Histopathology diagnosis was provided by consensus diagnosis from two study pathologists at BCH.

HRME images acquired in both studies were correlated with individual site biopsies, and resulting histopathology was used as the ground truth label for classification. From 1,627 participants, a total of N = 2,130 biopsy proven cervical images were acquired. Histopathology diagnostic categories include the following: negative, cervical intraepithelial neoplasia grade 1 (CIN 1), grade 2 (CIN 2), grade 3 (CIN 3), and invasive carcinoma.

Before undertaking algorithm development, the dataset was partitioned into training, validation, and test sets. Gradient-based optimization of model parameters was performed solely using the training set. Performance on the validation set was monitored during training to avoid overfitting. Finally, the test set was used to assess generalizability of a final model selected based on performance in the validation set. Partition assignments were generated on a per-patient using a stratified k-fold module to balance each fold by histopathology categories.⁸⁹ The data was initially split into five folds. Three of the five folds (60%) were assigned to the training set, one of the five (20%) assigned to validation, and the final fold (20%) assigned to test (**Figure 6-2**).

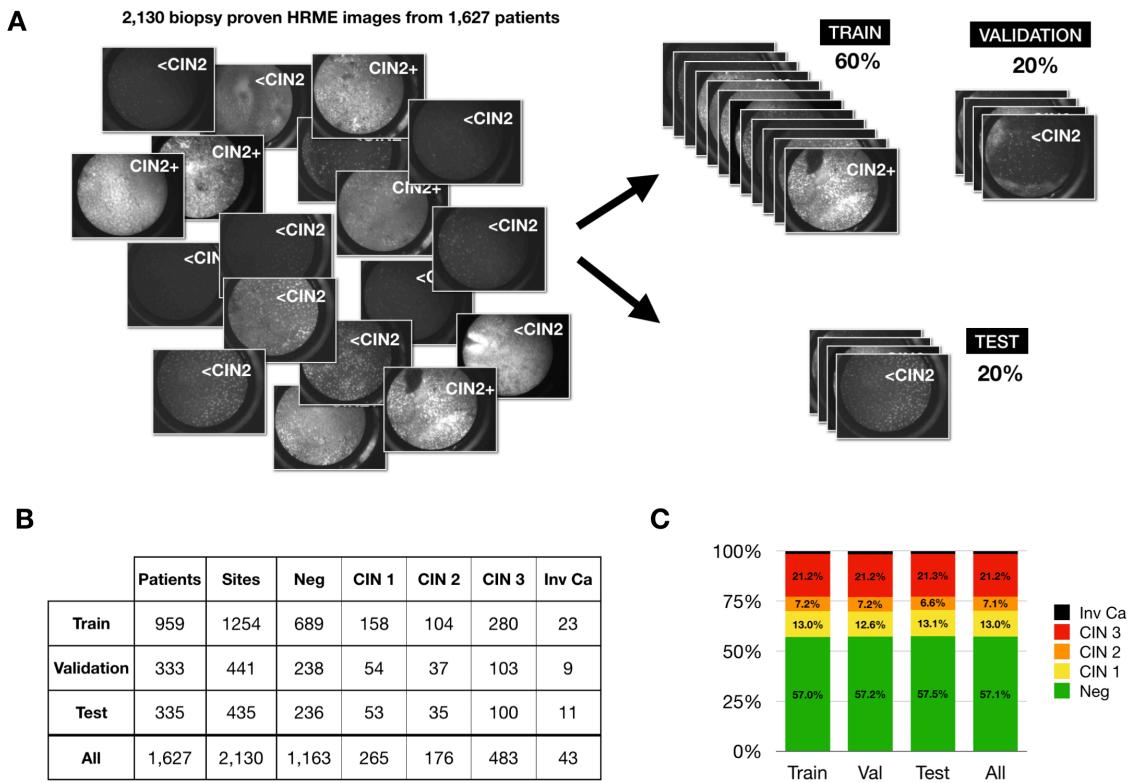


Figure 6-2 Partitioning of clinical study data into train, validation, and test sets. **A)** 2,130 biopsy proven images were split randomly by patient. Partitions were sampled such that each fold contained similar ratios of specific pathology categories. **B)** Resulting table containing total patients and sites. **C)** Stacked bar chart of pathology ratios in each partition.

6.2.2 Transfer learning using CNNs

At BCH a histopathology diagnosis of CIN 2 more severe (CIN 2+) is the clinical threshold used for deciding which patients should undergo preventative treatment (ablation or excision of precancerous tissue). As HRME would ideally provide real-time decision support for preventative treatment, a binary classification task was set up by grouping the histopathology labels as CIN 2+ (including CIN 2, CIN 3, and invasive carcinoma) vs <CIN 2 (including negative and CIN 1).

For classification using CNNs, I used a transfer learning approach starting with the ResNet18 network from He et al and the open source deep learning framework

PyTorch.^{24,90} The model is initialized with previously established parameters obtained from optimization of the model using 1.2 million images in the ImageNet dataset.⁹¹ The network was modified only slightly to perform binary classification by replacing the final dense layer and softmax layer with a new two node dense layer representing the two classes (<CIN 2 and CIN 2+).

Iterative gradient-based learning of model parameters was performed using cross entropy as the optimization loss function. To account for the difference in frequency of <CIN 2 vs CIN 2+ in the dataset, the loss function was weighted using the inverse frequency of each class. Minimization of this loss was performed for 50 epochs using a stochastic gradient descent (SGD) optimizer with momentum of 0.9 and a batch size of 4 images. The learning rate was decreased after every epoch with an initial learning rate as 0.001 and decay coefficient (gamma) of 0.95.

Mean cross entropy loss was calculated on training and validation sets at every epoch with results saved to a log file. Model parameters resulting in the lowest cross entropy loss on the validation set were saved. At the conclusion of training, best validation parameters were re-loaded from disk, and classification predictions generated for each image in the validation set. Validation predictions were used to compare diagnostic performance across all training runs. Because there is randomness involved in the optimization process, each optimization experiment was run in replicates of ten; average training and validation performance are reported across all replicates.

CNN optimization experiments were run via python Jupyter notebook interface using a high-performance desktop station composed of an Intel Core i7–8700K 3.7GHz 6-Core Processor, 32 GB of DDR4 RAM Memory, two Nvidia GeForce GTX 1080 Ti

GPUs, and 6 TB of solid disk memory. The desktop station ran a distribution of the Linux operating system (Ubuntu 16.04) and configured for GPU-accelerated computation using PyTorch.

6.2.3 Data preprocessing and augmentation transforms

The expected input of the ResNet architecture is an RGB image with dimensions of 224 x 224 pixels. The HRME utilizes a monochrome CCD sensor with dimensions of 1280 x 960 pixels (CMLN-13S2M-CS; Point Grey, Richmond, British Columbia, Canada), so HRME images were preprocessed to standardize them before input into the network (**Figure 6-3**). Raw HRME images contain a circular fiber bundle which may shift by small amounts based on mechanical insertion of the fiber between imaging sessions. Additionally, the magnification of the HRME system is such that the shift sometimes causes the edges of the fiber bundle to be cropped by the CCD sensor. In order to create a standardized image of the circular fiber bundle region of interest (fiber ROI) for input into the network, raw HRME images were preprocessed to center and crop the fiber ROI within each image. This step was performed once for each image using a circular Hough transform to fit a circle to a binary contour mask of the raw image. Based on this circular fit, a circular cropping was applied, and the image padded to center the circle on a 1024 x 1024 pixel image.

From fiber ROI images, two additional transformations were applied to achieve the input dimension of the network. The first is a simple resizing to 224 x 244, and the second is a resizing to 384 x 384 with subsequent cropping to 224 x 224 (**Figure 6-3**). The former preserves the circular structure of the fiber ROI image while the latter restricts the input image to only a square region of pixels within the fiber ROI. Performance of both of these approaches was compared during training.

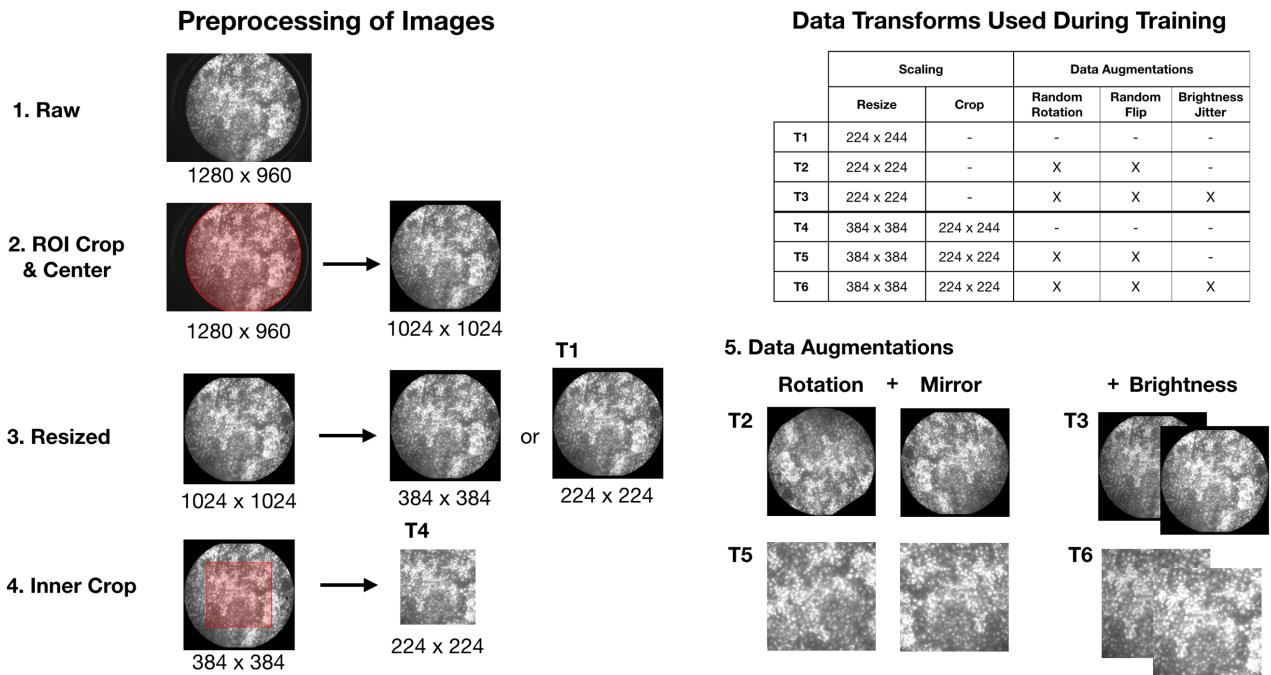


Figure 6-3 Image preprocessing steps and data augmentations utilized during training.

The aforementioned cropping and scaling transformations were applied to all image sets, but additional augmentations were applied to training images to prevent overfitting of the training images by randomly perturbing them without changing ground truth labeling. Three types of transformations supported by PyTorch were utilized in these experiments: random rotations (0 to 180 degrees), mirroring, and brightness adjustments (uniform scaling of brightness between 90 – 110%). Training was performed with and without data augmentations for a total of six conditions summarized in **Figure 6-3**.

6.2.4 Performance metrics and model selection

Mathematica 12.0 (Wolfram Research Inc., Champaign, Illinois, USA) was used to parse training log files and generate line plots of the training and validation loss over time. Training curves were averaged across the ten runs with the best performing models plotted as individual points on the averaged curve.

The metrics used to evaluate overall diagnostic performance on this dataset are area under the receiver operator characteristic curve (ROC AUC) and area under the precision recall curve (PR AUC). These operating curves were generated using the probability of CIN 2+ output from network as the quantitative parameter. An open source Mathematica module was used to generate ROC/PR curves and calculate the area under the curve (AUC) using the trapezoidal rule.⁹² The joint sum of ROC and PR AUC was used as the metric to select the best performing model. The best performing model was then evaluated on the test set using PyTorch.

Finally, GraphPad Prism 8.3 (GraphPad Software, La Jolla, California, USA) was used to generate column dot scatter plots and ROC curves to compare the best performing CNN model to the TCV algorithm on the validation and test sets. An open source R module was used for statistical comparison of ROC curves (DeLong's test) for both methods.⁹³

6.2.5 Deployment and speed benchmarks

The existing HRME instrumentation is controlled through a LabVIEW interface on a PC Tablet. The TCV algorithm is deployed within the system using a compiled MATLAB executable called by the LabVIEW application. I benchmarked the performance of image classification using CNNs without and without GPU acceleration on a commercial grade PC laptop (Surface Book 2, Microsoft). I compared the classification speed of ResNet to that of the TCV algorithm for single image classification as well as batched classification of 100 images. For ResNet, classification speed benchmarks were performed using timing functions in a python Jupyter notebook interface. For the TCV algorithm, timing functions were used inside MATLAB

2020a (The MathWorks Inc., Natick, Massachusetts, USA). Timing measurements were averaged across 20 replicates.

6.2.6 Source code availability

The source code for these experiments is available at the following URL:
<https://gitlab.com;bradyhunt/hrme-cervical-cnn-benchmarking>.

6.3 Results

Figure 6-4 shows averaged loss curves for all six experimental conditions. Without data augmentation (**Figure 6-4 A-B**), average validation loss (min-max) achieved was 0.591 (0.570-0.629). Furthermore, best validation model parameters were achieved on average after only 3.7 epochs of training, indicating quick onset of overfitting. The addition of rotational and mirroring transformations during training resulting in a 9.3% reduction in average validation loss (mean: 0.536, min-max: 0.513-0.561; Mann-Whitney U test: $p<0.001$) (**Figure 6-4 C-D**). These augmentations also significantly increased the average number of epochs until best validation model parameters were achieved (mean: 26.4; Mann-Whitney U test: $p<0.001$). The addition of brightness augmentations further reduced mean validation loss compared to rotation and mirroring alone, although not by a statistically significant margin (mean: 0.531 min-max: 0.517-0.549; Mann-Whitney U test: $p=0.21$) (**Figure 6-4 E-F**). For conditions using data augmentation (**Figure 6-4 C-F**), training on images that were cropped inside the fiber bundle ROI performed better than training on whole images (uncropped vs cropped mean loss: 0.540 vs 0.528; Mann-Whitney U test: $p<0.001$).

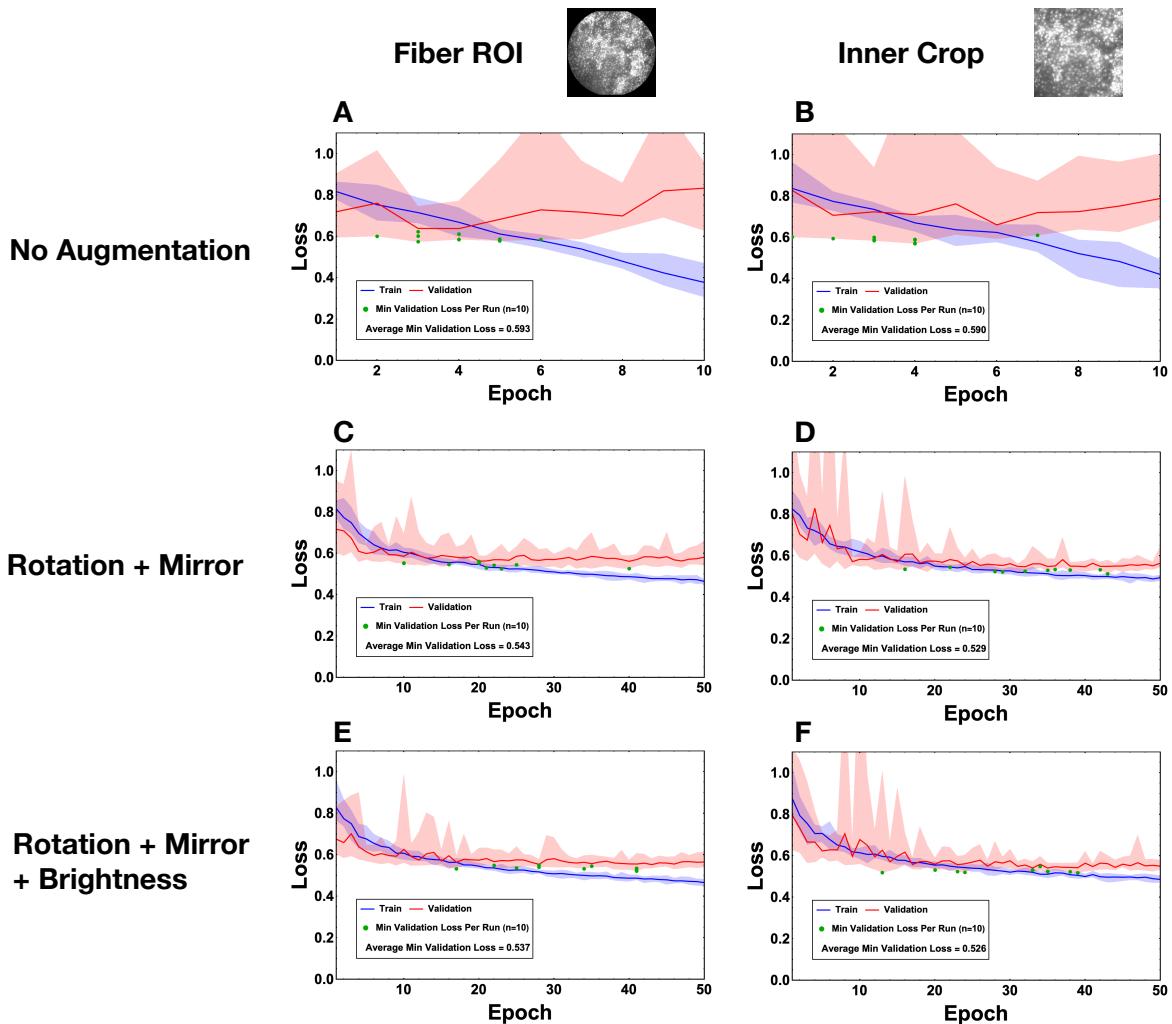


Figure 6-4: Training curves of convolutional neural network (ResNet18) with and without data augmentation. Line plots represent average training and validation loss over time. Shading on line plots represent minimum and maximum values across all replicates ($n=10$). Green points are the minimum validation loss achieved on each run.

Figure 6-5 contains scatter plots comparing the overall diagnostic performance using predictions on the validation dataset for all trained models ($n=60$, ten replicates in each condition). Points represent joint areas under the curve (AUC) for both receiver-operator characteristic (ROC) and precision-recall (PR). The best performing models (maximum area under ROC and PR curves) are located at the top right of the scatter plot. The best performing condition (T6) was inner cropping of the HRME fiber bundle ROI with all data augmentations, although the same set of transformations without

brightness adjustment (T5) did not differ by a significant margin (T5 vs T6, mean joint ROC and PR AUC: 1.431 vs 1.454; Mann-Whitney U test: $p=0.24$). The best performing model among all runs was from condition T5 and achieved 0.807 AUC ROC and 0.681 AUC PR respectively. This model was evaluated on the test set.

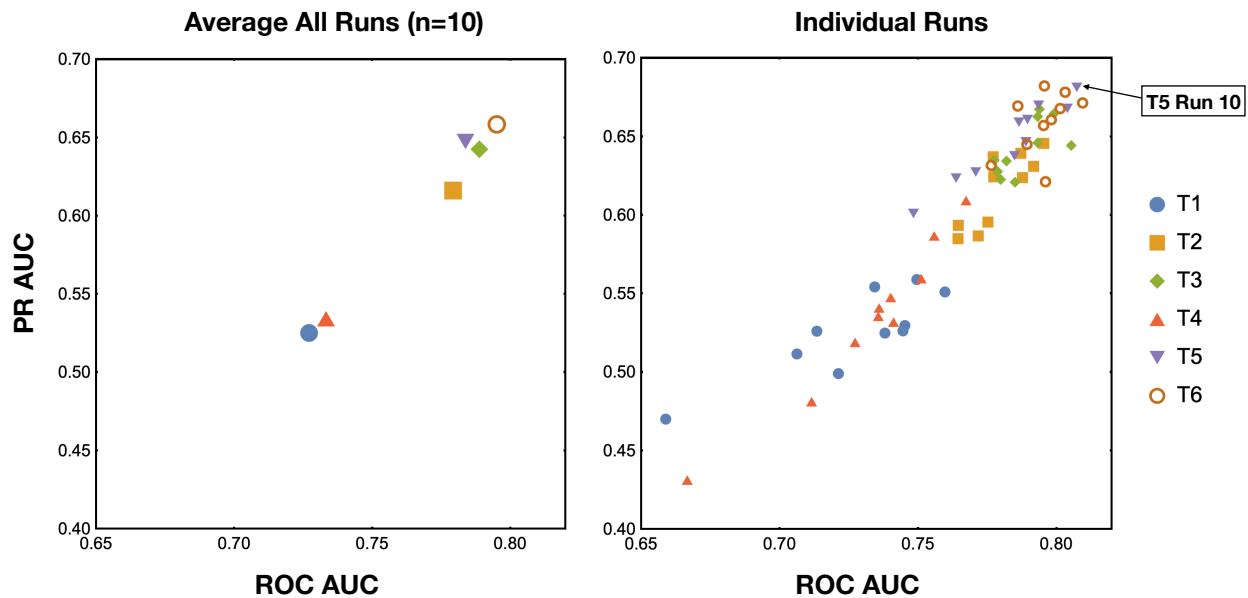


Figure 6-5: Comparison of validation performance across all training runs using joint receiver-operator characteristic (ROC) and precision-recall (PR) areas under the curve (AUC). Arrow indicates best performing model selected for evaluation on the test set.

Figure 6-6 compares diagnostic performance of the selected ResNet model to the TCV algorithm on the test set. Overall ROC AUC is compared as well as sensitivity (Sens) and specificity (Spec) at three different cutoffs on the curve. The AUC ROC of the ResNet algorithm outperformed the TCV algorithm by 6.6% (0.810 vs 0.760 AUC ROC; DeLong's test: $p=0.017$) (**Figure 6-6A**). Column dot scatter plots were used to visualize the distribution of predictive scores within each pathology category (**Figure 6-6 B-C**). Cutoffs used for Sens/Spec comparisons are visualized on the ROC curve as well as the column scatter plots. At the first cutoff, sensitivity was fixed at 90% and specificity

compared. At this cutoff, there was no significant difference in specificity for detection of CIN 2+ at this cutoff (TCV: 50.5% vs ResNet: 52.3%; p=0.47) (**Figure 6-6D**). At the second cutoff, balanced Sens/Spec was compared. There was no statistically significant difference in sensitivity (TCV: 71.9% vs ResNet: 73.3%; p=0.61) or specificity (TCV: 71.6% vs ResNet: 73.4%; p=0.61) (**Figure 6-6E**). At the third cutoff, specificity was fixed at 90% and sensitivity compared. At this cutoff, the sensitivity of ResNet exceeded that of the TCV algorithm by a statistically significant margin (TCV: 28.8% vs ResNet: 48.0%; p<0.001) (**Figure 6-6F**).

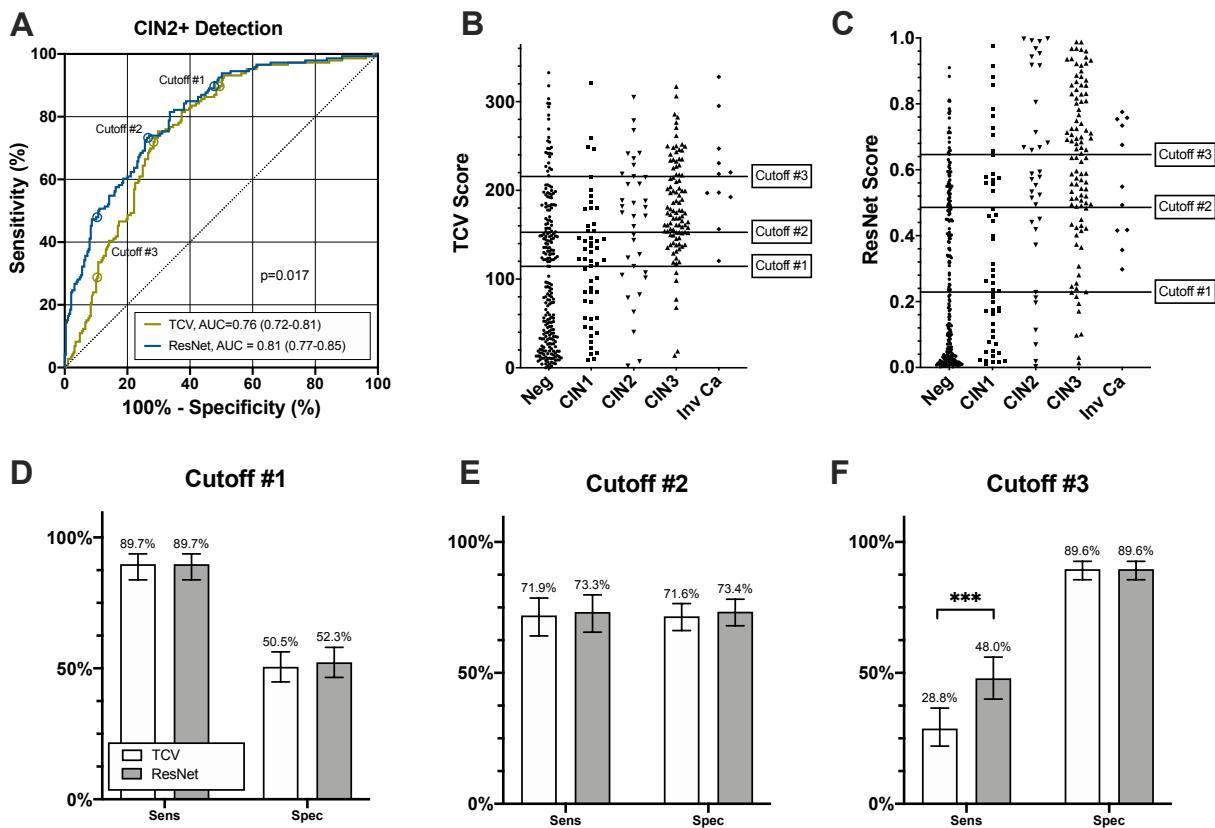


Figure 6-6: Diagnostic performance of ResNet model compared the existing traditional computer vision algorithm (TCV) on the test set. **A)** ROC curves for detection of CIN 2+ of both algorithms. Area under the curve (AUC) reported in figure legend with 95%CI in parentheses. Statistical comparison of AUC using DeLong's test. **B-C)** Column scatter plots of diagnostic scores by increasing pathology severity. Lines indicate location of three cutoffs for sensitivity and specificity comparisons. **D-F)** Bar plots of sensitivity (Sens) and specificity (Spec) at the three cutoffs indicated on the ROC curve and scatter plots. Error bars indicate 95% confidence intervals. Statistical comparisons were performed using McNemar's test (only significant values are shown).

Table 6-1 contains speed benchmarks for ResNet compared to the TCV algorithm on a consumer-grade laptop. As benchmarks were run by loading images from file storage, computation times are reported in three ways: time loading images into memory, time for classification, and total time. Average time to load images into memory ranged between 21-37 milliseconds (ms) per image, with MATLAB-based nuclei segmentation being slightly faster than python-based ResNet approach. However, classification speed was dramatically improved for ResNet compared to nuclei segmentation. Without GPU acceleration, single image classification for nuclei segmentation was 1050 ms compared to 129 ms for ResNet. GPU acceleration using ResNet further reduced single image classification to 11.7 ms. Finally, when using additional GPU memory to process images in a batch of 100 with ResNet, classification speed was reduced to 2.14 ms. Overall, the ResNet approach reduces total time for classification by approximately ~20x for single image classification (Nuclei Seg: 1071 ms vs ResNet: 49 ms) and batched classification time by up ~500x (Nuclei Seg: 1050 ms vs ResNet: 2.1 ms).

TCV vs ResNet	CPU vs GPU	Batch Size	Load Into Memory (ms)	Classification (ms)	Total (ms)
TCV	CPU	1	21.0	1050	1071
ResNet	CPU	1	30.6	129	169
ResNet	GPU	1	37.0	11.7	48.7
ResNet	GPU	100	33.9*	2.14*	36.0*

*Reported times are normalized per image

Table 6-1: Speed benchmark of ResNet compared to traditional computer vision (TCV) on a consumer-grade laptop.

6.4 Discussion

This study evaluated baseline performance of an existing convolutional neural network for diagnostic classification of HRME images. The direct advantages of this approach for automated detection of cervical dysplasia are 1) a statistically significant increase in area under the receiver operator characteristic curve measured using a hold-out test set, 2) dramatically faster image classification times in deployment. Of these two benefits, the classification time reduction is highly promising as it enables truly real-time diagnostic evolution at speeds greater than 20 frames per second (50ms per frame). More broadly, this study demonstrates a machine learning approach to HRME image classification which requires less manual supervision than previous methods and establishes baseline performance benchmarks for the approach. The dataset utilized in this study is the largest library of HRME images with biopsy results and should be a valuable resource for ongoing development of automated image analysis methods for HRME. This work utilizes publicly available open-source software libraries and is generalizable to additional HRME image classification tasks.

While the machine learning approach presented has some promising advantages over previous methods, there are also important limitations to this method and considerations for future work. Machine learning depends on the availability of sufficient training data to support learning of generalizable features that can accurately classify future examples. This is especially challenging in the case of medical imaging where acquiring representative images of various disease conditions is very costly and time consuming. In this study, the only supervision provided for optimization of model parameters was the binarized histopathology classification label. Oakden-Rayner and colleagues have pointed out that such coarse-grained labeling schemes (normal vs

abnormal) can lead to a phenomenon they call *hidden stratification* or simply unrecognized subsets within the data.⁹⁴ Hidden stratification can lead to falsely reassuring aggregate performance metrics which are dominated by the most frequent subsets within the data. In the case *in vivo* microscopy of cervical tissues, images of columnar and metaplasia tissues are an example of one stratum within this dataset that is challenging for automated image analysis approaches using HRME, but there may be others. The binary ResNet classifier presented does not sufficiently address these challenging subsets in the dataset.

Oakden-Rayner et al proposed three approaches to dealing with hidden stratification: 1) schema completion, 2) error auditing, and 3) algorithmic measure.⁹⁴ While all three should be very beneficial, schema completion seems particularly relevant for histopathology prediction. Schema completion involves creating a complete set of labeled subclasses within the dataset. Such subclasses would be particularly helpful for images that are negative for cervical intraepithelial neoplasia and give rise to false positives with existing HRME image analysis approaches. I have initiated an effort to identify helpful subclasses by reviewing the textual writing of the pathology reports, but it will take additional time and discussion with clinical collaborators to arrive at a schema that can better support machine learning approaches to HRME image classification.

In summary, this study presents a dataset which can be utilized for developing improved diagnostic classifiers for HRME images, establishes performance benchmarks on the dataset using an existing convolutional neural network, and provides important considerations for ongoing development of HRME image analysis approaches using machine learning.

7. FASTHRME: A HIGH FRAME RATE MICROENDOSCOPE CAPABLE OF ROBUST FREEHAND MOSAICKING

The contents of this chapter have been prepared with assistance from Jackson Coole, Alex Kortum, David Brenes, Ruchika Mitbander, Imran Vohra, Jennifer Carns, Richard Schwarz, and Rebecca Richards-Kortum and will be submitted for publication.

7.1 Abstract

High-resolution microendoscopy (HRME) is a low-cost, real-time alternative to tissue biopsy. Current strategies incorporating HRME are based on acquisition and analysis of still images. This approach requires acquisition of multiple images to assess regions of interest larger than the field of view of the probe ($\sim 0.5 \text{ mm}^2$) and relies on clinical judgment for probe placement. Mosaicking approaches to increase the field of view of HRME have previously been demonstrated, but the frame rate limits of previous systems made them impractical for freehand video mosaicking. Here, we present the *FastHRME*, an improved microendoscope system incorporating a high frame-rate CMOS sensor and optical probe holder to enable high-speed, high quality scanning of tissues. We developed a mosaicking approach using *FastHRME* videos and assessed mosaic registration accuracy, image sharpness, and tolerable translational speeds during freehand acquisition. *FastHRME* tolerates *in vivo* probe translation of up to 15 millimeters per second while preserving high image quality and accurate mosaicking.

7.2 Introduction

The high-resolution microendoscope (HRME) is a fiber-bundle imaging device which can characterize size, shape, and distribution of epithelial cell nuclei *in vivo*, aiding in detection of dysplasia and cancer in mucosal tissues.^{34,68} Images may be interpreted by clinician or analyzed quantitatively by software algorithms to generate

an automated diagnostic prediction in real time.^{38,95,96} Anatomic sites include oral cavity^{36,97}; cervix^{79,81}; esophagus^{13,37}; and colon.⁹⁸ Applications include aiding in the evaluation of previously identified lesions for dysplasia or cancer; aiding in biopsy guidance; and potentially aiding in surgical margin determination. Current strategies incorporating HRME are based on acquisition and analysis of still images. This approach works well for evaluation of localized lesions in which clinician judgment can guide probe placement. For biopsy guidance, still imaging has more limited utility due to the time and effort required to explore a tissue area larger than a few square millimeters by sequential probe placement and imaging. In the case of surgical margin assessment, the ability to smoothly move the probe across the border of a clinically identified lesion while imaging continuously is needed to determine the extent of the lesion. The ability to do freehand video mosaicking with real-time automated image analysis would dramatically expand the utility of the HRME for biopsy guidance during surveillance and for intraoperative surgical guidance applications.

Freehand mosaicking using fiber bundle microscopes is limited by a number of factors. First, camera exposure times need to be short enough to avoid motion blur. Second, the ability to mechanically translate the optical probe while maintaining good contact with the specimen is also necessary. If these two factors cannot be satisfied, video mosaicking approaches will be of limited use as the image quality will be poor. Beyond the aforementioned factors, the primary determinants of tolerable translational speeds are frame rate and single frame field-of-view (FOV). Acquisition frame rate needs to be sufficiently fast to obtain significant overlap between frames, and larger FOV optical probes are preferable to smaller ones because they can be translated over larger distances while preserving sufficient frame-to-frame overlap. While there are some trade-offs that can be made between these factors, reasonable translational speeds

of all fiber bundle imaging systems will in some way be ultimately limited by one or more of them.

There have been relatively few studies which have focused on freehand video mosaicking using fiber bundle microscopy. In 2012, Bedard et al presented a real-time video mosaicking approach using HRME instrumentation with a 4 μm lateral resolution, 800 μm circular FOV, and acquisition frame rate of 11 frames per second (FPS).⁹⁹ Bedard et al demonstrated both *in vivo* and *ex vivo* mosaics ranging from 2-5 mm². Maximum probe translation speed of the system was 1 millimeter per second (mm/s), limited primarily by a 10 ms exposure and 11 FPS frame rate. In 2015, Hughes et al presented a system which achieved even faster probe translation speeds through significantly increased acquisition frame rates.¹⁰⁰ Their line-scanning confocal system had a lateral resolution of 1 μm , 240 μm circular FOV, and achieved a frame rate of 120 FPS. They assessed average mosaic lengths within a 3 min video of freehand motion across porcine colon during a simulated surgical procedure. They were able to register 86% of frames into smaller scale mosaics that were on average 2.3 mm in length (~0.6 mm²). They demonstrated successful mosaicking at translational speeds up to 5 mm/s, and were limited primarily by the small FOV of their system.

Here, we present the *FastHRME* (fast acquisition with stable translation), an improved microendoscope system incorporating a high frame-rate CMOS sensor and optical probe holder to enable higher speed scanning of tissues and more robust freehand mosaicking.

7.3 Materials and methods

7.3.1 Optical assembly

Figure 7-1 contains an optical diagram (**Figure 7-1A**) and solidworks illustration (**Figure 7-1B**) of the *FastHRME*. The optical assembly is an adaptation of previous designs which utilize an optical fiber (FIGH-30-850N; Fujikura), LED illumination source, objective lens (RMS10X; Olympus), dichroic mirror, and filter set suitable for fluorescence imaging of proflavine stained specimens.^{34,68,96} The primary change to the system is the integration of a high frame-rate CMOS sensor (BFS-U3-04S2M-CS; FLIR Systems Inc., Richmond, BC, Canada). This 0.4 megapixel sensor uses a global shutter and has a maximum framerate of 522 frames per second (FPS). In order to account for the smaller sensor format, the magnification of the system was reduced using a shorter focal length tube lens (AC254-100-A-ML; Thorlabs Inc., Newton, NJ, USA). The optical assembly is mounted inside a 3D printed enclosure and the CMOS sensor connected to a laptop (Surfacebook 2; Microsoft, Redmond, Wash, USA) via USB connection (**Figure 7-1C**). The LED is powered via the USB connection through an LED driver circuit

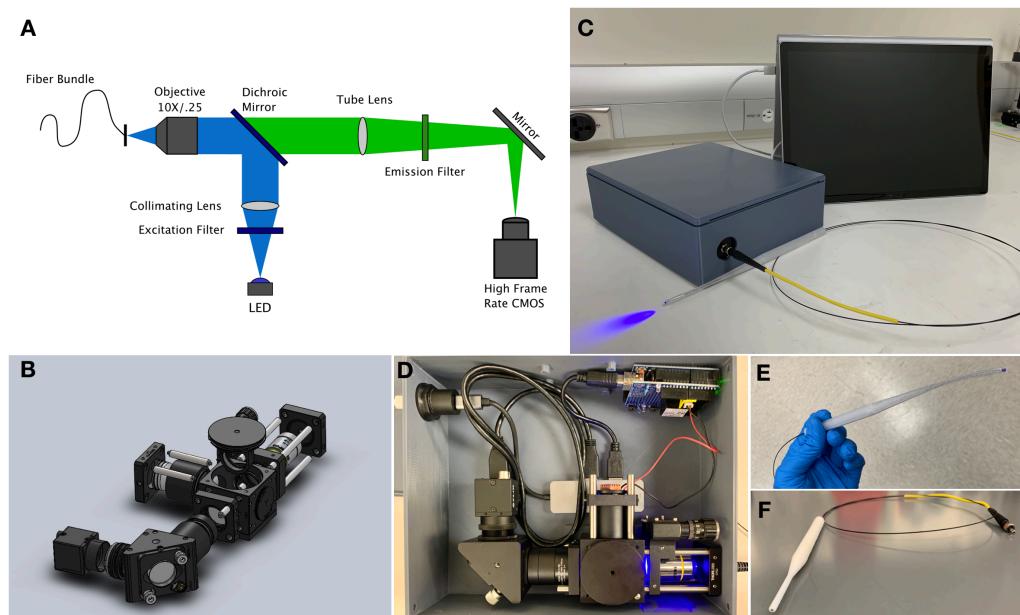


Figure 7-1: Schematic of *FastHRME*. **A)** Diagram of optical components. **B)** Solidworks illustration of optical components. **C)** Photograph of assembled device connected to the laptop. **D)** Photograph of internal view of assembled device. **E-F)** 3D printed probe holders used to improve probe stability in optical target imaging (**E**) and *in vivo* imaging (**F**) experiments.

mounted on an Arduino Uno microcontroller (**Figure 7-1D**). When driven at full brightness, the system delivers ~4 mW of blue light at the distal tip of the optical fiber. The LED brightness can be modulated using a potentiometer on the driver circuit.

7.3.2 Optical fiber holder

Two optical fiber holder designs were created in Solidworks 2019 (Solidworks Corp., Waltham, MA, USA) and fabricated using 3D printing. The first design has a 200 mm long handle with a subtle curve tapering towards a 2 mm diameter tip (**Figure 7-1E**). This design was printed in two pieces out of PC-ISO polymer using a Fortus 450mc printer (Stratasys Ltd., Eden Prairie, MN, USA) and then assembled onto an optical fiber using epoxy adhere the two pieces together. The second design is linear cylinder with a 100 mm long handle tapering to a 4 mm diameter tip (**Figure 7-1F**). This design was printed as a monolithic piece out of white resin (FLGPWH04) on a stereolithography printer (Form 2; Formlabs, Somerville, MA, USA). The second design contains channels to insert the optical fiber and secure it using two nylon tipped set screws. The first design was utilized during optical target imaging experiments, and the second design was utilized during *in vivo* imaging experiments.

7.3.3 Optical target imaging

Before conducting *in vivo* assessments of the system, we first acquired high frame-rate videos while translating the probe by freehand across an optical target with a known template geometry. A plastic guide was fabricated using vector graphics software (Adobe Illustrator 2020) and a CO₂ laser cutter (Universal Laser Systems VLS3.60) to precisely cut the template out of an acetate film. The plastic guide was then adhered on top of lens paper that had been dyed using a yellow highlighter. The lens paper was backed with packing foam to provide some mechanical flexibility as the

probe was pressed against the template during translation. Video files were recorded as using SpinView (FLIR Systems Inc., Richmond, BC, Canada) with a fixed exposure, gain, and framerate.

7.3.4 In vivo imaging

In vivo imaging was performed at two sites in the oral cavity of a healthy adult volunteer–mucosa of the lower lip and lateral border of the tongue. Proflavine at a concentration of 0.01% was applied topically to each site prior to imaging. *FastHRME* video was again acquired using SpinView. For visual reference of the probe movement in the oral cavity, a circular disk was laser cut out of water-activated kraft paper tape (KGPTI-0200; Alanson Products, Miami, FL, USA) and adhered to the tissue. Widefield video of the oral cavity imaging was acquired using a smartphone system with magnifying optics (EVA 3 Plus, MobileODT, Tel Aviv, Israel). Data collection was approved by the Institutional Review Board of Rice University (ID#: IRB-FY2016-172) and written informed consent was obtained from the participant.

7.3.5 Video processing and mosaic construction

Figure 7-2 shows the sequence of image processing steps for construction of mosaics from videos acquired of optical targets and oral tissues. Software to perform these processing steps using a video file as input was developed using Mathematica 12.1 (Wolfram Research Inc., Champaign, IL, USA). Video files were processed on a MacBook Pro running macOS 10.14 (Apple, Cupertino, CA, USA). Videos were reviewed to identify a start and end timepoint for processing each video. Preprocessing of video frames prior to mosaic construction was performed in five steps. First, raw frames between the specified time points were extracted as 8-bit grayscale images. Next,

high-frequency intensity patterns imposed by the fiber bundle were removed by down sampling the image by a factor of 2.5 using Lanczos method to 288×216 pixels. Then, intensities across all frames were averaged to compute a mean intensity image for each video. The mean intensity image was binarized to create a mask for the region within the video sequence corresponding to the fiber bundle. Additionally, the mean intensity mask was used to identify any statically dim pixels and interpolate them using neighboring pixels. The last preprocessing step was brightness equalization using the mean intensity image to reduce boundary effects when overlaying images during mosaic construction.

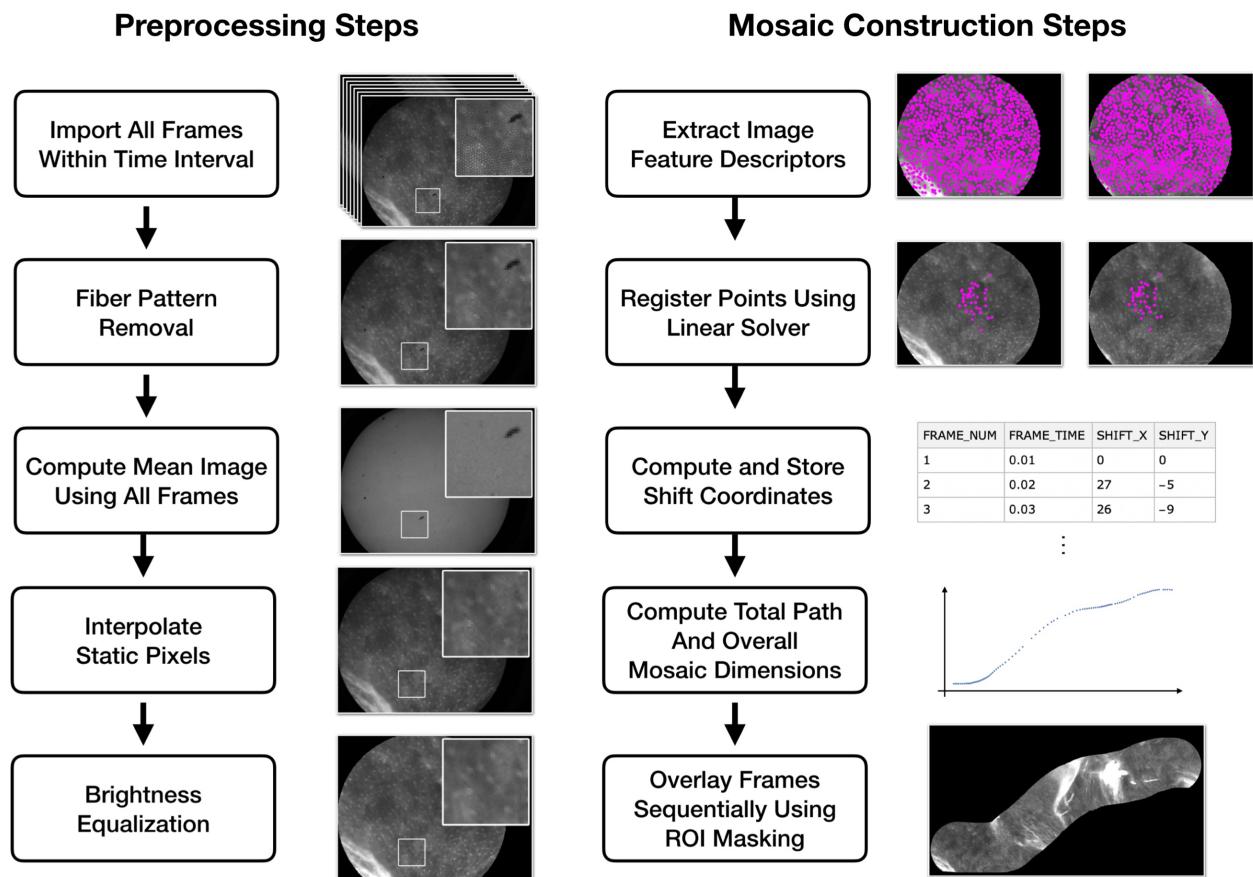


Figure 7-2: Steps for preprocessing and mosaic construction using *FastHRME* video.

The first two steps in the mosaicking procedure were to extract a 2D feature descriptor and then register image keypoints between frames. For these steps, a Mathematica function called '*ImageCorrespondingPoints*' was used to determine keypoints contained in both images. The descriptor used was a combination of feature extractors available in Mathematica and included KAZE and SURF features.^{101,102} Pixel locations of matched keypoints were then used to solve for a linear transformation and compute the pixel shift between the two images. The cumulative sum of the pixel shifts was used to determine the total dimensions of the mosaic. A blank image of the required dimensions was initialized, and frames inserted sequentially based on the registered shifts.

7.3.5 Quantitative metrics for assessing video mosaic quality

Errors in mosaic registration can occur when probe movement is too fast or when there the probe is in poor contact with the specimen. Tissue deformations, low signal, and oversaturation can also make registration challenging. Therefore, three quantitative metrics were computed frame by frame during mosaic construction to assess frame registration accuracy, image sharpness, and probe speed during acquisition. These metrics were exported at each time point and plotted as time series for each video mosaic presented.

Mosaic registration accuracy was assessed using root-mean-square error (RMSE) of the overlapping regions for each frame inserted into the mosaic. This metric was used to reject frames with high registration error. Two error thresholds for rejection were determined by computing the registration error between randomly shuffled frames in the video sequence and taking the 1st and 10th percentiles of that distribution as possible rejection thresholds. The 1st percentile threshold is a more stringent rejection threshold

whereas the 10th percentile is more lenient. These thresholds and number of frames rejected are plotted on the registration error line plot for each mosaic.

In order to assess loss of contact or motion blurring during acquisition, image sharpness was also assessed. The image sharpness was computed as the standard deviation in image intensities after high-frequency filtering using a 5x5 pixel Laplacian of Gaussian kernel. Because oversaturated regions can heavily skew this sharpness metric, only non-saturated pixel regions were used in computing sharpness.

Lastly, probe speed was quantified by converting the Euclidean distance of the pixel shift to millimeters, and then multiplying by the acquisition frequency to obtain velocity in millimeters per second. At time points where frames were rejected for high registration error, linear interpolation was used to determine probe speed. As frame-to-frame velocity can be quite volatile, a moving average for probe speed was also calculated using 50 frame intervals.

7.3.7 Graphical visualizations

Mosaic images were exported from the mosaicking software in JPEG format. Annotation of these files to add scalebars and illustrate magnified insets at various points in the mosaic were created using Pixelmator Pro 1.6 (Pixelmator Team Ltd., Vilnius, Lithuania). Mosaic images were rotated, cropped, and exported using Pixelmator. No brightness or contrast adjustments were made. Time series plots for registration accuracy, image quality, and probe speed were generated in GraphPad Prism 8.4 (GraphPad Software, La Jolla, California, USA).

7.4 Results

Figure 7-3 shows the mosaic constructed from a video sequence obtained by freehand imaging of a linear optical target. The target was rectangular in shape with dimensions of 50 x 2 mm. The optical probe was moved in single linear motion from one side of the target to the other. The video sequence used for mosaic construction contains 855 frames and was acquired in 4.275 seconds (200 FPS). Camera exposure and gain were fixed at 100 μ s and 0 dB respectively. The resulting mosaic image has dimensions of 50.0 x 2.0 mm and a filled in area of 33 mm². Average registration error for the video sequence was 0.028 (min-max: 0.012-0.045) and all frames were well below the 1st percentile rejection threshold. Average image sharpness for frames in the sequence was 0.035 (min-max: 0.028-0.042). Average probe speed during acquisition was 11 mm/s and the maximum probe speed registered was 51 mm/s.

Figure 7-4 shows the mosaic constructed from a video sequence obtained by freehand imaging of a curved optical target. The target shape was cursive letters of the word 'Rice' with dimensions of 53.5 x 22 mm. The optical probe was moved in single motion from top left of the target by tracing out the letters. The video sequence used for mosaic construction contains 4,124 frames and was acquired in 20.62 seconds (200 FPS). Camera acquisition settings were the same as the previous target (100 μ s exposure no gain). The resulting mosaic image has dimensions of 53.6 x 22.0 mm and a filled in area of 132 mm². Average registration error for the video sequence was 0.027 (min-max: 0.006-0.171). One frame was above the 10th percentile rejection threshold and excluded from the mosaic. Average image sharpness for frames in the sequence was 0.034 (min-max: 0.013-0.044), but some motion blur was observed in frames with the lowest sharpness. Average probe speed during acquisition was 9.2 mm/s and the maximum probe speed registered was 111 mm/s. Excluding frames in the bottom quartile for

image sharpness (>0.031 sharpness), the maximum probe speed registered was 95 mm/s.

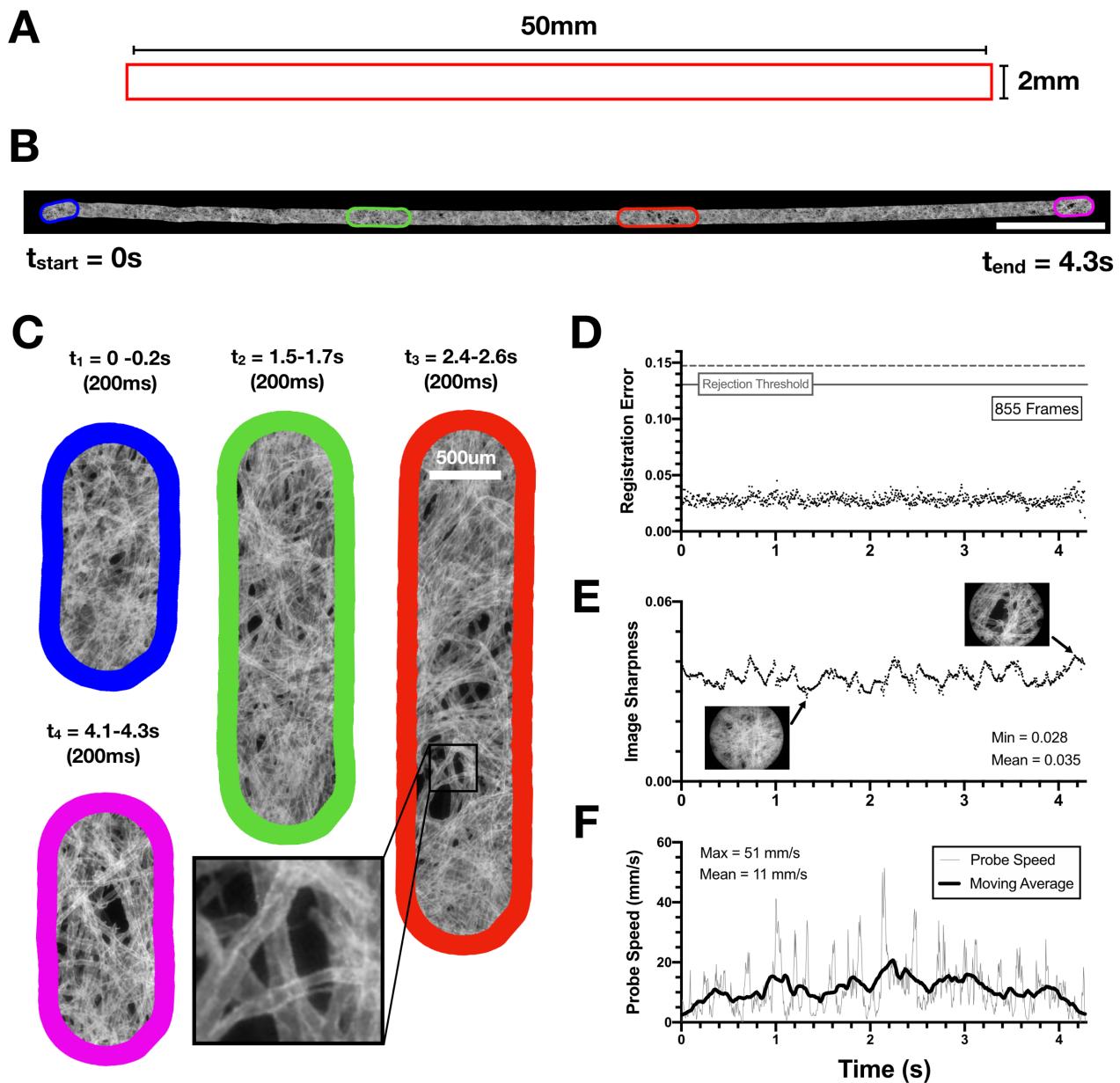


Figure 7-3: Freehand imaging of a linear target on lens paper. **A)** Laser cut template dimensions. **B)** Resulting mosaic constructed from 855 frames. Scale bar is 5mm. The mosaic image has dimensions of 50 x 2 mm and a filled area of 33 mm². Colored boxes overlaid on mosaic represent 200ms time intervals at four time points. **C)** Magnified view of color insets with time point indicated. **D)** Registration error for each frame. 1st percentile and 10th percentile error rejection thresholds are indicated by solid and dashed lines respectively. No frames were rejected in the video sequence. **E)** Image sharpness for each frame. Images of min/max sharpness frames are overlaid on plot. **F)** Estimated translational speed of probe tip during video acquisition. Moving average is calculated using 0.25 second intervals.

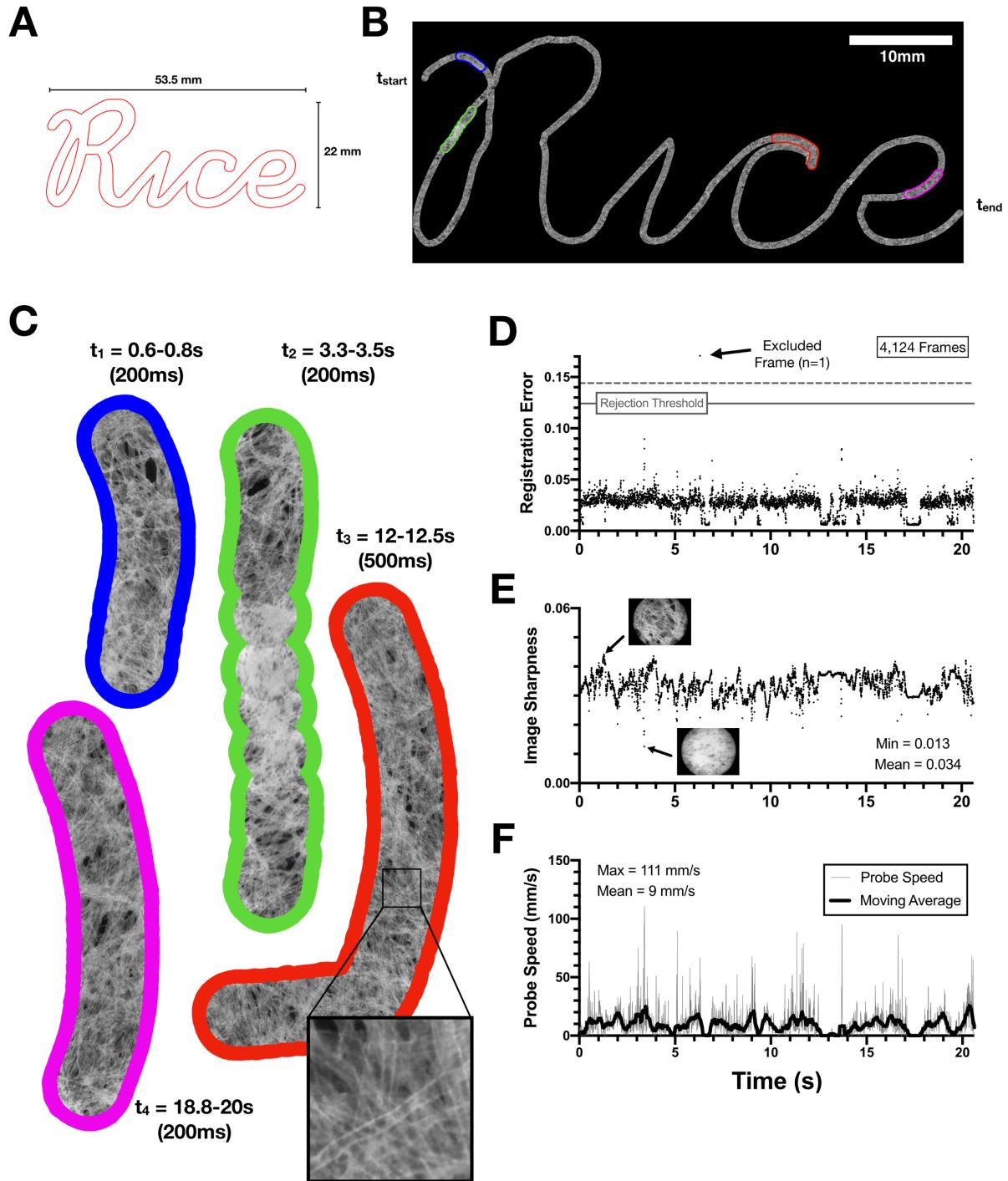


Figure 7-4: Freehand imaging of a curved target on lens paper. **A)** Laser cut template dimensions. **B)** Resulting mosaic constructed from 4,124 frames. The mosaic image has dimensions of 53.6×23.5 mm and a filled area of 123 mm^2 . Colored boxes overlaid on mosaic indicate inset magnified regions. **C)** Magnified view of color insets with time point indicated. **D)** Registration error for each frame. 1st percentile and 10th percentile error rejection thresholds are indicated by solid and dashed lines respectively. One frame was rejected for high registration error. **E)** Image sharpness for each frame. Images of min/max sharpness frames are overlaid on plot. **F)** Estimated translational speed of probe tip during video acquisition. Moving average is calculated using 0.25 second intervals.

For *in vivo* imaging of proflavine stained tissues, exposure and gain were increased to 1ms and 18 dB respectively. Due to the increased exposure, the maximum achievable framerate over USB connection on the laptop was ~100 FPS. **Figure 7-5** shows the mosaic constructed from a video sequence obtained by freehand imaging of an 8mm circular region of oral mucosa. Cell nuclei are visible as bright dots within the magnified image insets (**Figure 7-5C**). The optical probe was moved in spiral motion from the outside of the circle to the center. The video sequence used for mosaic construction contains 1,114 frames and was acquired in 11.17 seconds (99.7 FPS). The resulting mosaic image has dimensions of 8.1 x 8.1 mm and a filled in area of 21 mm². Average registration error for the video sequence was 0.028 (min-max: 0.016-0.099). Two frames were above the 1st percentile rejection threshold and excluded from the mosaic. Average image sharpness for frames in the sequence was 0.021 (min-max: 0.013-0.025). Comparing frames with the lowest and highest sharpness, both are in focus and free from motion artifacts. The primary difference is that the lower sharpness frames are from regions of the tissue with lower overall brightness. Average probe speed during acquisition was 2.8 mm/s and the maximum probe speed registered was 15 mm/s.

Figure 7-6 shows the mosaic constructed from a video sequence obtained by freehand imaging of the tongue. The optical probe was moved in linear motion from ventral to dorsal sides of the tongue, crossing over the boundary where lingual papillae are present. The video sequence used for mosaic construction contains 268 frames and was acquired in 2.85 seconds (94.0 FPS). The resulting mosaic image has dimensions of 3.7 x 14.7 mm and a filled in area of 10 mm². Large bright structures in the latter half of the video sequence are likely keratinized filiform papillae. Average registration error for the video sequence was 0.040 (min-max: 0.018-0.357). Five frames were above the 10th percentile rejection threshold and excluded from the mosaic. Average image

sharpness for frames in the sequence was 0.017 (min-max: 0.009-0.022), and motion blur was observed in frames with low sharpness. Average probe speed during acquisition was 5.7 mm/s and the maximum probe speed registered was 33 mm/s. A sharp movement due to the probe slipping was observed in the widefield video at time 2.66 seconds, corresponding to the timeframe in which high registration errors were observed. It is likely that probe speed exceeded 33 mm/s in some of the five frames with highest registration error.

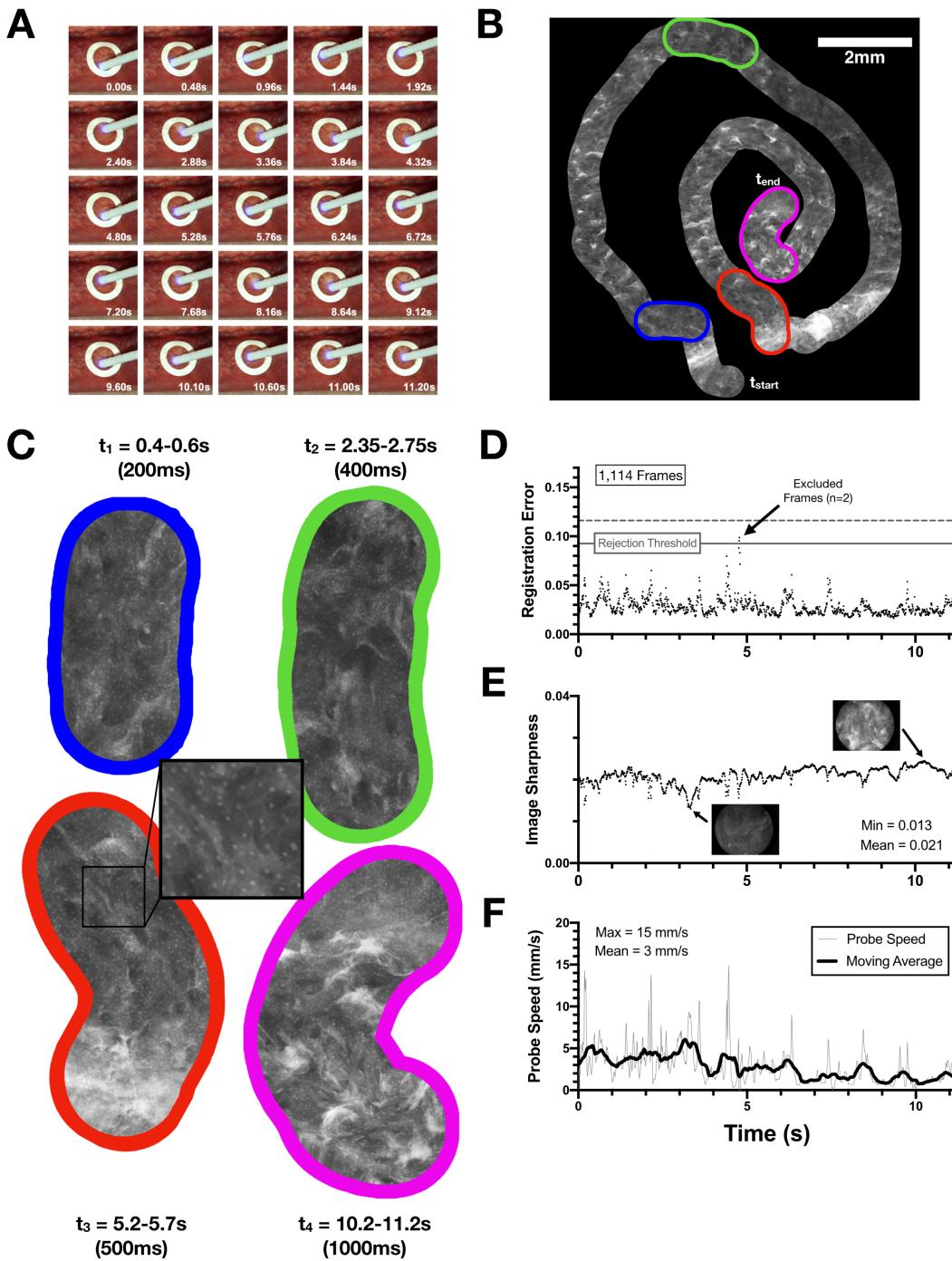


Figure 7-5: Freehand imaging of a circular target on oral mucosa. **A)** Probe placement within an 8mm laser cut sticker template. The probe was moved in a spiral motion from outside to inside of the circular region. **B)** Resulting mosaic constructed from 1,114 frames. The mosaic image has dimensions of 8.1 x 8.1 mm and a filled area of 21 mm². Colored boxes overlaid on mosaic indicate inset magnified regions. **C)** Magnified view of color insets with time points indicated. **D)** Registration error for each frame. 1st percentile and 10th percentile error rejection thresholds are indicated by solid and dashed lines respectively. Two frames were rejected for high registration error. **E)** Image sharpness for each frame. Images of min/max sharpness frames are overlaid on plot. **F)** Estimated translational speed of probe tip during video acquisition. Moving average is calculated using 0.5 second intervals.

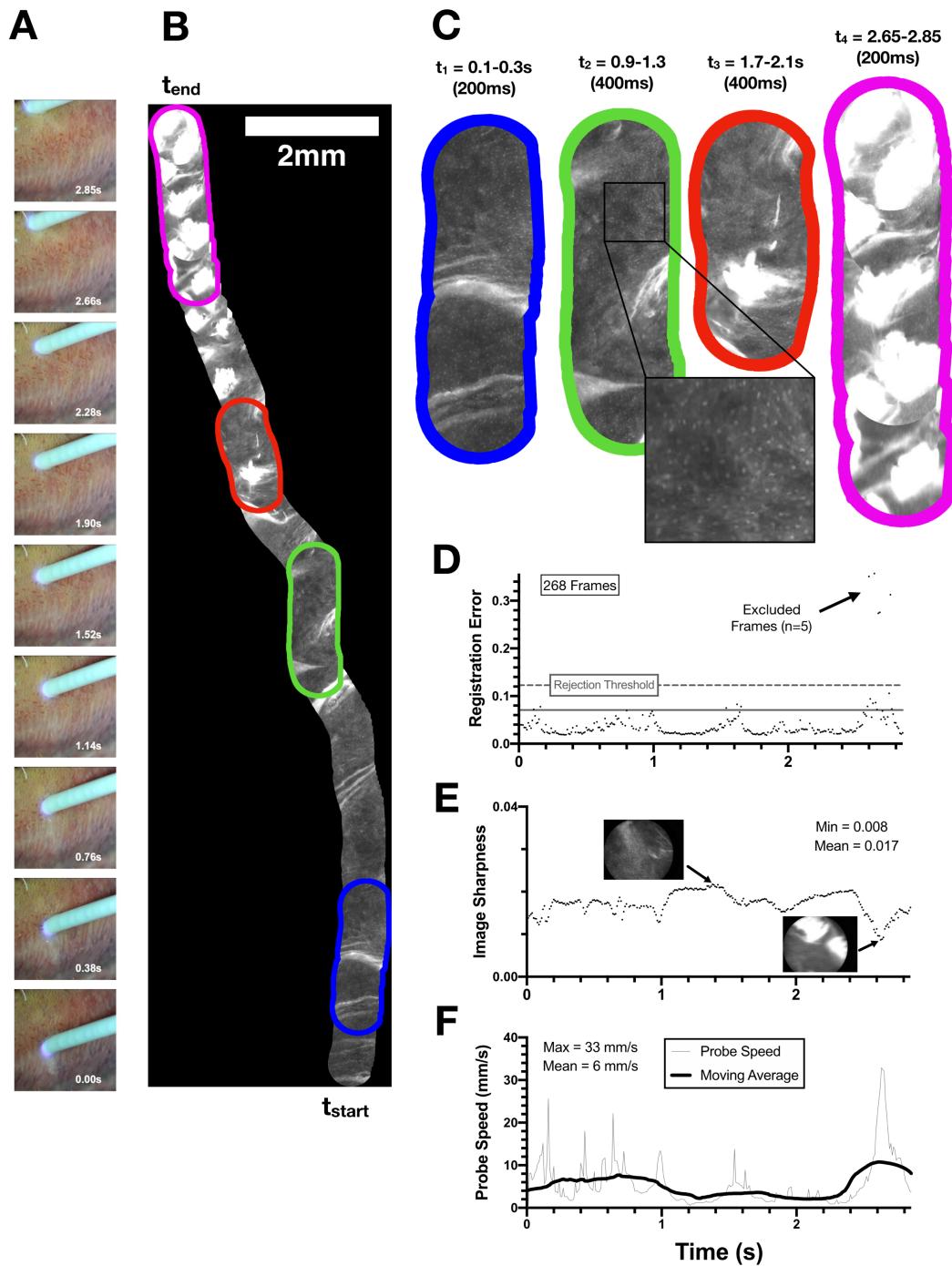


Figure 7-6: Freehand imaging of tongue boundary. **A)** Probe placement on tongue. The probe was moved in a linear motion from ventral to dorsal sides of the tongue crossing over the boundary where lingual papillae are present. **B)** Resulting mosaic constructed from 268 frames. The mosaic image has dimensions of 3.7×14.7 mm and a filled area of 10 mm^2 . Colored boxes overlaid on mosaic indicate inset magnified regions. **C)** Magnified view of color insets with time points indicated. **D)** Registration error for each frame. 1st percentile and 10th percentile error rejection thresholds are indicated by solid and dashed lines respectively. The 10th percentile threshold was used, and five frames were rejected from the video sequence. **E)** Image sharpness for each frame. Images of min/max sharpness frames are overlaid on plot. **F)** Estimated translational speed of probe tip during video acquisition. Moving average is calculated using 0.5 second intervals.

7.5 Discussion

In this manuscript, we present an improved widefield microendoscopy system capable of imaging larger surface areas at high-speed. This is achieved through incorporation of a high-frame rate CMOS into the optical design as well as an optical probe holder to support stable freehand translation. During optical target and *in vivo* imaging experiments, we obtained frame rates between 100-200 FPS over USB connection to a laptop. These frame rates supported accurate mosaic registration at speeds of at least 50 mm / s in optical targets and 15 mm / s in oral tissues. Qualitatively, the mosaics presented accurately reproduce template geometries with minimal artifacts introduced from the mosaicking process. Quantitatively, we defined a rejection error threshold for successful mosaic registrations by comparing actual registration errors to that of a random distribution from the same video sequence. For video sequences ranging from 3 to 20 seconds long, we observed very low frame rejection rates (between 0.1% and 2%). *In vivo* mosaics presented contain clearly visible nuclei structures and tissue boundaries with total areas that are 20 to 40 times larger than a single image FOV.

*Fast*HRME achieves high frame rate acquisition while maintaining the low cost and complexity of previous HRME systems. The resolution of the system is limited by the 4 μm spacing of the optical fiber core spacing. Numerous clinical studies involving HRME have demonstrated this is sufficient for real-time diagnostic image analysis approaches using nuclei morphology assessment. Therefore, the potential diagnostic capability of the *Fast*HRME is at least as good as previous HRME systems and may be better due to mitigation of motion blur and ability to survey larger areas of tissue.

The mosaicking software used in this work employs raw video as input and can process a 1000 frame mosaic ($\sim 20 \text{ mm}^2$) in under 5 minutes. Although this may not be

fast enough for clinical translation, it can be improved upon in future work. The bulk of this computation time (~80%) is taken by the feature extraction and pixel shift registration, which takes approximately 250 ms per frame. This aspect of the algorithm could be improved by utilizing image features that are less computationally expensive as well as parallel computing architectures.

Many have noted the challenge of tissue deformations for *in vivo* mosaicking approaches.^{100,103-106} Prior work has demonstrated freehand mosaics ranging from 2-6 mm².^{99,100,105,106} Our system achieves mosaics that are over 100 mm² in optical targets and 20 mm² in oral tissues. We were able to obtain high-quality, large-scale mosaics using only linear registrations due to increased frame-to-frame overlap from the high frame rate. Still, future developments could utilize this increased sampling to perform more sophisticated mosaic registration methods, including super-resolution approaches.¹⁰⁷⁻¹⁰⁹

Due to USB bandwidth limitations, frame rates achieved in these experiments were lower than the 522 FPS rating specified by the sensor vendor. Further optimization to achieve the full frame rate of the camera would increase scanning speeds by a factor of 5. This would require additional computation time to perform mosaic registrations and downstream analyses, but may be an adequate tradeoff in very high-speed imaging applications.

While this work did not conduct a clinical evaluation of the *FastHRME* system, it provides a proof-of-principle that should have significant impact on clinical strategies utilizing HRME. High-speed imaging can significantly reduce the time required to survey tissues, better guide biopsy selection, and strengthen correlation with subsequent tissue biopsies. *FastHRME* is extremely well suited for integration with other new technological developments such as automated probe tracking and deep-

learning based image analysis algorithms. Multimodal imaging approaches that incorporate these advances would provide clinicians with powerful and versatile new diagnostic imaging tools to aid in the detection, surveillance, and treatment of dysplasia and cancer.

8. CONCLUSION

8.1 Summary of results

This dissertation is composed a series of clinical and technical studies conducted to evaluate and then improve a low-cost, high-resolution microendoscope for real-time diagnostic evaluation of cervical lesions.

In the first study, a van was adapted into a mobile diagnostic unit to perform diagnostic examination by colposcopy and HRME close to home. Exams were performed in over 38 rural communities, and a 37% relative increase in diagnostic completion rate was observed for patients seen at the mobile diagnostic van compared to those seen at the central hospital (mobile van: 87%, central hospital: 64%, $p < 0.001$).

The second study evaluated the safety of HRME imaging by following patients and assessing rates of disease progression between those exposed to proflavine ($n=232$) and matched control group not exposed ($n=160$). Mean duration of follow-up was 18.7 and 20.1 months for the proflavine-exposed and controls groups, respectively. There were no significant differences in disease progression between the proflavine exposed and control groups, suggesting that exposure to dilute proflavine does not increase the risk of cervical precancer and cancer progression.

The third study was a large-scale, prospective evaluation of the diagnostic performance of the HRME in 1,600 screen positive women. The HRME automated image analysis algorithm was found to have 95.6% sensitivity (95%CI: 92.8–97.4%) and 56.6% specificity (95%CI: 53.7–59.4%) for detection of high-grade cervical abnormalities. By comparison, diagnostic performance of the standard-of-care method, colposcopy, had 96.2% sensitivity (HRME vs colposcopy, McNemar's test: $p=0.81$) and 58.7%

specificity ($p=0.18$). Areas for improvement of HRME were identified including: 1) better algorithms to address challenges in analyzing images of columnar/metaplasia tissue and 2) improved optical instrumentation to ensure good spatial correlation between the location imaged and the site of the subsequent tissue biopsy.

In the fourth study, machine learning was utilized to develop an improved HRME classifier based on an existing convolutional neural network (ResNet). A total of 2,130 HRME cervical images with biopsy results were partitioned into training, validation, and test sets. The ResNet classifier improved area under the receiver operator characteristic curve for detection of high-grade cervical dysplasia by 6.6% (AUC 0.81 vs 0.76, DeLong's test: $p=0.02$). Additionally, the ResNet algorithm was deployed on a consumer grade laptop with GPU acceleration, reducing single image classification speed to less than 50 milliseconds per image (previously 1000 ms per image).

In the last study, I present the *FastHRME*, an improved microendoscope system capable of robust *in vivo* mosaicking. This is accomplished by incorporating a high-frame rate CMOS sensor and a 3D printed optical fiber holder for stable freehand probe movement. The *FastHRME* increases acquisition frame rate from 18 frames per second up to 200 frames per second, and supports probe translation up to 15 millimeters per second while maintaining high image quality and accurate video mosaicking. A method for video mosaicking using *FastHRME* was developed, and *in vivo* mosaics presented contain clearly visible nuclei structures and tissue boundaries with total areas that are 20 to 40 times larger than a single HRME image FOV.

8.2 Future work

The cumulative studies contained in this dissertation provide strong evidence for the potential of HRME to improve cervical cancer prevention in low-resource settings. In order for that potential to be reached, additional work is needed.

From a technology perspective, the *FastHRME* overcomes a significant hurdle for HRME since its initial use in cervical lesion assessment—the limited FOV of the optical probe. The ability to translate the probe smoothly across tissue while acquiring high-quality video can enable surveillance of larger tissue areas and mitigate biases in clinical judgment of probe placement. However just acquiring high-speed video doesn't immediately translate to improved usability for the HRME at the POC. Further development of computational approaches to utilize these videos is needed, including methods for faster mosaic construction and diagnostic evaluation of videos.

This work suggests that machine learning can provide a way to develop faster and more accurate diagnostic classifiers. However, even if *FastHRME* video can be acquired of the entire cervix, histopathology labels cannot be obtained for the entire cervix. Therefore, there is also need for development of strategies to effectively annotate resulting *FastHRME* videos to enable them to be utilized for machine learning. Initial work with *FastHRME* has demonstrated the utility of acquiring simultaneous wide-field video to give spatial context to microendoscopy video. The combination of *FastHRME* with a mobile colposcope has high potential as a multimodal cervical imaging system. Multimodal cervical imaging could utilize automated probe tracking and annotations of clinical impressions on the wide-field video to provide clinically meaningful annotations to corresponding *FastHRME* video segments. This approach can

dramatically increase the amount of training data for machine learning approaches, and would ensure better generalizability of machine learning classifiers.

In conclusion, the clinical studies and resulting technology developments contained in this work represent a sustained effort over many years of collaboration with wonderful colleagues at Barretos Cancer Hospital and The University of Texas MD Anderson Cancer Center. Participation in this clinical research collaboration has given me a global health engineering perspective that I could not have obtained any other way. My sincere hope is that such collaborations will continue and that the diagnostic imaging technologies we continue to develop will assist in the elimination of cervical cancer.

9. REFERENCES

1. Ferlay J, Colombet M, Soerjomataram I, et al. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int J Cancer* 2018.
2. Torre LA, Siegel RL, Ward EM, Jemal A. Global Cancer Incidence and Mortality Rates and Trends--An Update. *Cancer Epidemiol Biomarkers Prev* 2016; **25**(1): 16-27.
3. Ferlay J, Ervik M, Lam F, et al. Global Cancer Observatory: Cancer Today. 2018. <https://gco.iarc.fr/today> (accessed January 19 2019).
4. Schiffman M, Doorbar J, Wentzensen N, et al. Carcinogenic human papillomavirus infection. *Nature Reviews Disease Primers* 2016; **2**: 20.
5. Guan P, Howell-Jones R, Li N, et al. Human papillomavirus types in 115,789 HPV-positive women: a meta-analysis from cervical infection to cancer. *International journal of cancer* 2012; **131**(10): 2349-59.
6. Schiffman M, Castle PE, Jeronimo J, Rodriguez AC, Wacholder S. Human papillomavirus and cervical cancer. *The Lancet* 2007; **370**(9590): 890-907.
7. Sankaranarayanan R. Screening for cancer in low- and middle-income countries. *Ann Glob Health* 2014; **80**(5): 412-7.
8. Kuroki LM, Bergeron LM, Gao F, Thaker PH, Massad LS. See-and-Treat Loop Electrosurgical Excision Procedure for High-Grade Cervical Cytology: Are We Overtreating? *Journal of Lower Genital Tract Disease* 2016; **20**(3): 247-51.
9. Nghiem VT, Davies KR, Beck JR, Follen M, Cantor SB. Overtreatment and Cost-Effectiveness of the See-and-Treat Strategy for Managing Cervical Precancer. *Cancer Epidemiology Biomarkers & Prevention* 2016; **25**(5): 807-14.
10. Quinn MK, Bubi TC, Pierce MC, Kayembe MK, Ramogola-Masire D, Richards-Kortum R. High-resolution microendoscopy for the detection of cervical neoplasia in low-resource settings. *PLoS One* 2012; **7**(9): e44924.
11. Pierce MC, Guan Y, Quinn MK, et al. A pilot study of low-cost, high-resolution microendoscopy as a tool for identifying women with cervical precancer. *Cancer Prev Res (Phila)* 2012; **5**(11): 1273-9.
12. Grant BD, Fregnani JH, Possati Resende JC, et al. High-resolution microendoscopy: a point-of-care diagnostic for cervical dysplasia in low-resource settings. *Eur J Cancer Prev* 2017; **26**(1): 63-70.
13. Quang T, Schwarz RA, Dawsey SM, et al. A tablet-interfaced high-resolution microendoscope with automated image interpretation for real-time evaluation of esophageal squamous cell neoplasia. *Gastrointest Endosc* 2016; **84**(5): 834-41.
14. Grant BD. A low-cost platform for cervical cancer diagnosis: Rice University; 2016.
15. Santesso N, Mustafa RA, Schunemann HJ, et al. World Health Organization Guidelines for treatment of cervical intraepithelial neoplasia 2-3 and screen-and-treat strategies to prevent cervical cancer. *Int J Gynaecol Obstet* 2016; **132**(3): 252-8.
16. Qiao L, Li B, Long M, Wang X, Wang AR, Zhang GN. Accuracy of visual inspection with acetic acid and with Lugol's iodine for cervical cancer screening: Meta-analysis. *Journal of Obstetrics and Gynaecology Research* 2015; **41**(9): 1313-25.
17. Asiedu MN, Simhal AK, Chaudhary U, et al. Development of algorithms for automated detection of cervical pre-cancers with a low-cost, point-of-care, Pocket Colposcope. *bioRxiv* 2018: 324541.

18. Hu L, Bell D, Antani S, et al. An observational study of deep learning and automated evaluation of cervical images for cancer screening. *JNCI: Journal of the National Cancer Institute* 2019; **111**(9): 923-32.
19. Machine learning—Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Machine_learning&oldid=879230491 (accessed January 11, 2019).
20. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015; **521**(7553): 436-44.
21. LeCun Y, Boser B, Denker JS, et al. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation* 1989; **1**(4): 541-51.
22. Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 1998; **86**(11): 2278-324.
23. Abadi M, Agarwal A, Barham P, et al. TensorFlow: large-scale machine learning on heterogeneous systems. Software available from tensorflow.org. 2015. URL <https://www.tensorflow.org> 2015.
24. Paszke A, Gross S, Massa F, et al. PyTorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems*; 2019; 2019. p. 8024-35.
25. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin* 2011; **61**(2): 69-90.
26. Brasil. Estimativas 2016-2017: Incidência de câncer no Brasil.
27. Silveira CE, Silva TB, Fregnani JH, et al. Digital photography in skin cancer screening by mobile units in remote areas of Brazil. *BMC Dermatol* 2014; **14**: 19.
28. Mauad EC, Silva TB, Latorre MR, et al. Opportunistic screening for skin cancer using a mobile unit in Brazil. *BMC Dermatol* 2011; **11**: 12.
29. Mauad EC, Nicolau SM, Gomes UA, et al. Can mobile units improve the strategies for cervical cancer prevention? *Diagn Cytopathol* 2010; **38**(10): 727-30.
30. Lorenzi AT, Fregnani JH, Possati-Resende JC, et al. Can the careHPV test performed in mobile units replace cytology for screening in rural and remote areas? *Cancer Cytopathol* 2016; **124**(8): 581-8.
31. Haikel RL, Jr., Mauad EC, Silva TB, et al. Mammography-based screening program: preliminary results from a first 2-year round in a Brazilian region using mobile and fixed units. *BMC Womens Health* 2012; **12**: 32.
32. Fregnani JH, Scapulatempo C, Haikel RL, Jr., et al. Could alarmingly high rates of negative diagnoses in remote rural areas be minimized with liquid-based cytology? Preliminary results from the RODEO Study Team. *Acta Cytol* 2013; **57**(1): 69-74.
33. Faria EF, Carvalhal GF, Vieira RA, Silva TB, Mauad EC, Carvalho AL. Program for prostate cancer screening using a mobile unit: results from Brazil. *Urology* 2010; **76**(5): 1052-7.
34. Muldoon TJ, Pierce MC, Nida DL, Williams MD, Gillenwater A, Richards-Kortum R. Subcellular-resolution molecular imaging within living tissue by fiber microendoscopy. *Opt Express* 2007; **15**(25): 16413-23.
35. Pierce M, Yu D, Richards-Kortum R. High-resolution fiber-optic microendoscopy for in situ cellular imaging. *J Vis Exp* 2011; (47).
36. Pierce MC, Schwarz RA, Bhattacharjee VS, et al. Accuracy of in vivo multimodal optical imaging for detection of oral neoplasia. *Cancer Prev Res (Phila)* 2012; **5**(6): 801-9.
37. Protano MA, Xu H, Wang G, et al. Low-Cost High-Resolution Microendoscopy for the Detection of Esophageal Squamous Cell Neoplasia: An International Trial. *Gastroenterology* 2015; **149**(2): 321-9.

38. Shin D, Protano MA, Polydorides AD, et al. Quantitative analysis of high-resolution microendoscopic images for diagnosis of esophageal squamous cell carcinoma. *Clin Gastroenterol Hepatol* 2015; **13**(2): 272-9.e2.
39. Kurman RJ, Carcangiu ML, Harrington CS, Young RH. WHO Classification of Tumours of Female Reproductive Organs.
40. World Health Organization. Treatment of cervical intraepithelial neoplasia 2-3 and adenocarcinoma in situ: cryotherapy, large loop excision of the transformation zone, and cold knife conization. 2014.
http://apps.who.int/iris/bitstream/10665/104174/1/9789241506779_eng.pdf.
41. Donner A, Klar N. Design and Analysis of Cluster Randomization Trials in Health Research.
42. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009; **42**(2): 377-81.
43. R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing.
44. Stevenson M. CRAN package - epiR: Tools for the Analysis of Epidemiological Data.; 2017.
45. Fay MP. Two-sided exact tests and matching confidence intervals for discrete data. *R journal* 2010; **2**(1): 53-8.
46. Pimple S, Mishra G, Shastri S. Global strategies for cervical cancer prevention. *Curr Opin Obstet Gynecol* 2016; **28**(1): 4-10.
47. Mitchell G, Buttle G. Proflavine in closed wounds. *The Lancet* 1943; **242**(6276): 749.
48. Wainwright M. Acridine—a neglected antibacterial chromophore. *Journal of Antimicrobial Chemotherapy* 2001; **47**(1): 1-13.
49. Watret L, White R. Surgical wound management: the role of dressings. *Nursing Standard* 2001; **15**(44): 59-69.
50. Rahman G, Adigun I, Yusuf I, Ofoegbu C. Wound dressing where there is limitation of choice. *Nigerian Journal of Surgical Research* 2006; **8**(3-4): 151-4.
51. Janssen PA, Selwood BL, Dobson SR, Peacock D, Thiessen PN. To dye or not to dye: a randomized, clinical trial of a triple dye / alcohol regime versus dry cord care. *Pediatrics* 2003; **111**(1): 15-20.
52. McConnell TP, Lee CW, Couillard M, Sherrill WW. Trends in umbilical cord care: scientific evidence for practice. *Newborn and Infant Nursing Reviews* 2004; **4**(4): 211-22.
53. Kiesslich R, Burg J, Vieth M, et al. Confocal laser endoscopy for diagnosing intraepithelial neoplasias and colorectal cancer in vivo. *Gastroenterology* 2004; **127**(3): 706-13.
54. Polglase AL, McLaren WJ, Skinner SA, Kiesslich R, Neurath MF, Delaney PM. A fluorescence confocal endomicroscope for in vivo microscopy of the upper- and the lower-GI tract. *Gastrointest Endosc* 2005; **62**(5): 686-95.
55. Leong RW, Nguyen NQ, Meredith CG, et al. In vivo confocal endomicroscopy in the diagnosis and evaluation of celiac disease. *Gastroenterology* 2008; **135**(6): 1870-6.
56. Günther U, Daum S, Heller F, et al. Diagnostic value of confocal endomicroscopy in celiac disease. *Endoscopy* 2010; **42**(03): 197-202.
57. Li C-Q, Yu T, Zuo X-L, et al. Effects on confocal laser endomicroscopy image quality by different acriflavine concentrations. *Journal of Interventional Gastroenterology* 2011; **1**(2): 59-63.

58. Fuchs FS, Zirlik S, Hildner K, Schubert J, Vieth M, Neurath MF. Confocal laser endomicroscopy for diagnosing lung cancer in vivo. *European Respiratory Journal* 2013; **41**(6): 1401-8.
59. Pierce MC, Javier DJ, Richards-Kortum R. Optical contrast agents and imaging systems for detection and diagnosis of cancer. *Int J Cancer* 2008; **123**(9): 1979-90.
60. National Toxicology Program. Bioassay of proflavine for possible carcinogenicity. *Natl Cancer Inst Carcinog Tech Rep Ser* 1977; **5**: 1-103.
61. International Agency for Research on Cancer. IARC monographs on the evaluation of carcinogenic risk of chemicals to man: some miscellaneous pharmaceutical substances. Lyon, France; 1977. p. 31-8.
62. International Agency for Research on Cancer. IARC monographs on the evaluation of carcinogenic risk of chemicals to humans: some pharmaceutical drugs. Lyon, France; 1980. p. 195-210.
63. Obstoy B, Salaun M, Veresezan L, et al. Safety and performance analysis of acriflavine and methylene blue for in vivo imaging of precancerous lesions using fibered confocal fluorescence microscopy (FCFM): an experimental study. *BMC Pulmonary Medicine* 2015; **15**(1): 30.
64. Lee K, Zhang H, Qian DZ, Rey S, Liu JO, Semenza GL. Acriflavine inhibits HIF-1 dimerization, tumor growth, and vascularization. *Proc Natl Acad Sci U S A* 2009; **106**(42): 17910-5.
65. Pepin G, Nejad C, Thomas BJ, et al. Activation of cGAS-dependent antiviral responses by DNA intercalating agents. *Nucleic Acids Research* 2017; **45**(1): 198-205.
66. Marselos M, Vainio H. Carcinogenic properties of pharmaceutical agents evaluated in the IARC Monographs programme. *Carcinogenesis* 1991; **12**(10): 1751-66.
67. International Agency for Research on Cancer. IARC Monographs- Classifications. 2018. http://monographs.iarc.fr/ENG/Classification/latest_classif.php (accessed January 12, 2018).
68. Pierce M, Yu D, Richards-Kortum R. High-resolution Fiber-optic Microendoscopy for in situ Cellular Imaging. *JoVE* 2011; (47): e2306.
69. Therneau TM. A Package for Survival Analysis in S. 2015. <https://CRAN.R-project.org/package=survival>.
70. Harrington DP, Fleming TR. A Class of Rank Test Procedures for Censored Survival Data. *Biometrika* 1982; **69**(3): 553-66.
71. Östör AG. Natural history of cervical intraepithelial neoplasia: a critical review. *International Journal of Gynecological Pathology* 1993; **12**(2): 186.
72. Mills AM, Paquette C, Castle PE, Stoler MH. Risk stratification by p16 immunostaining of CIN1 biopsies: a retrospective study of patients from the quadrivalent HPV vaccine trials. *Am J Surg Pathol* 2015; **39**(5): 611-7.
73. Arbyn M, Weiderpass E, Bruni L, et al. Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis. *Lancet Glob Health* 2020; **8**(2): e191-e203.
74. World Health Organization. Accelerating the elimination of cervical cancer as a global public health problem: World Health Organization. Regional Office for South-East Asia, 2019.
75. Austad K, Chary A, Xocop SM, et al. Barriers to Cervical Cancer Screening and the Cervical Cancer Care Continuum in Rural Guatemala: A Mixed-Method Analysis. *J Glob Oncol* 2018; **4**: 1-10.
76. Denny L, De Sousa M, Kuhn L, Pollack A, Wright TC. Cervical cancer prevention—A paradigm shift? *Gynecologic oncology* 2005; **99**(3): S12.

77. Kuhn L, Saidu R, Boa R, et al. Clinical evaluation of modifications to a human papillomavirus assay to optimise its utility for cervical cancer screening in low-resource settings: a diagnostic accuracy study. *The Lancet Global Health* 2020; **8**(2): e296-e304.
78. Brasil. Estimativa 2020: Incidência de câncer no Brasil.
79. Hunt B, Fregnani JHT, Schwarz RA, et al. Diagnosing cervical neoplasia in rural Brazil using a mobile van equipped with in vivo microscopy: A cluster-randomized community trial. *Cancer Prev Res (Phila)* 2018.
80. Liu J, Loewke NO, Mandella MJ, et al. Real-time pathology through in vivo microscopy. *Stud Health Technol Inform* 2013; **185**: 235-64.
81. Parra SG, Rodriguez AM, Cherry KD, et al. Low-cost, high-resolution imaging for detecting cervical precancer in medically-underserved areas of Texas. *Gynecologic oncology* 2019; **154**(3): 558-64.
82. Pierce M, Yu D, Richards-Kortum R. High-resolution fiber-optic microendoscopy for in situ cellular imaging. *JoVE (Journal of Visualized Experiments)* 2011; (47): e2306.
83. Pantano N, Hunt B, Schwarz RA, et al. Is Proflavine Exposure Associated with Disease Progression in Women with Cervical Dysplasia? A Brief Report. *Photochemistry and Photobiology* 2018; **94**(6): 1308-13.
84. Agresti A, Coull BA. Approximate is better than "exact" for interval estimation of binomial proportions. *The American Statistician* 1998; **52**(2): 119-26.
85. Scheffer M, Cassenote A, Guilloux AGA, Biancarelli A, Miotto BA, Mainardi GM. Demografia médica no Brasil 2018. 2018.
86. Adesina A, Chumba D, Nelson AM, et al. Improvement of pathology in sub-Saharan Africa. *The lancet oncology* 2013; **14**(4): e152-e7.
87. Wu ES, Jeronimo J, Feldman S. Barriers and challenges to treatment alternatives for early-stage cervical cancer in lower-resource settings. *Journal of global oncology* 2017; **3**(5): 572-82.
88. Long J, Parker HE, Ehrlich K, Tanner MG, Dhaliwal K, Mills B. Frugal filtering optical lenses for point-of-care diagnostics. *Biomedical Optics Express* 2020; **11**(4): 1864-75.
89. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python. *Journal of machine learning research* 2011; **12**(Oct): 2825-30.
90. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition; 2016; 2016. p. 770-8.
91. Russakovsky O, Deng J, Su H, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision* 2015; **115**(3): 211-52.
92. Antonov A. MathematicaForPrediction, GitHub repository. 2019. <https://github.com/antononcube/MathematicaForPrediction>.
93. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC bioinformatics* 2011; **12**(1): 77.
94. Oakden-Rayner L, Dunnmon J, Carneiro G, Ré C. Hidden stratification causes clinically meaningful failures in machine learning for medical imaging. *arXiv preprint arXiv:190912475* 2019.
95. Muldoon TJ, Thekkek N, Roblyer DM, et al. Evaluation of quantitative image analysis criteria for the high-resolution microendoscopic detection of neoplasia in Barrett's esophagus. *Journal of biomedical optics* 2010; **15**(2): 026027.
96. Muldoon TJ, Roblyer D, Williams MD, Stepanek VM, Richards-Kortum R, Gillenwater AM. Noninvasive imaging of oral neoplasia with a high-resolution fiber-optic microendoscope. *Head & neck* 2012; **34**(3): 305-12.

97. Yang EC, Vohra IS, Badaoui H, et al. Development of an integrated multimodal optical imaging system with real-time image analysis for the evaluation of oral premalignant lesions. *Journal of biomedical optics* 2019; **24**(2): 025003.
98. Tang Y, Polydorides AD, Anandasabapathy S, Richards-Kortum RR. Quantitative analysis of in vivo high-resolution microendoscopic images for the detection of neoplastic colorectal polyps. *Journal of biomedical optics* 2018; **23**(11): 116003.
99. Bedard N, Quang T, Schmeler K, Richards-Kortum R, Tkaczyk TS. Real-time video mosaicing with a high-resolution microendoscope. *Biomed Opt Express* 2012; **3**(10): 2428-35.
100. Hughes M, Yang G-Z. High speed, line-scanning, fiber bundle fluorescence confocal endomicroscopy for improved mosaicking. *Biomedical optics express* 2015; **6**(4): 1241-52.
101. Alcantarilla PF, Bartoli A, Davison AJ. KAZE features. European Conference on Computer Vision; 2012: Springer; 2012. p. 214-27.
102. Bay H, Ess A, Tuytelaars T, Van Gool L. Speeded-up robust features (SURF). *Computer vision and image understanding* 2008; **110**(3): 346-59.
103. Vercauteren T, Meining A, Lacombe F, Perchant A. Real time autonomous video image registration for endomicroscopy: fighting the compromises. Three-Dimensional and Multidimensional Microscopy: Image Acquisition and Processing XV; 2008: International Society for Optics and Photonics; 2008. p. 68610C.
104. Erden MS, Rosa B, Szewczyk J, Morel G. Understanding soft-tissue behavior for application to microlaparoscopic surface scan. *IEEE transactions on biomedical engineering* 2012; **60**(4): 1059-68.
105. Thrapp AD, Hughes MR. Automatic motion compensation for structured illumination endomicroscopy using a flexible fiber bundle. *Journal of Biomedical Optics* 2020; **25**(2): 026501.
106. Yin C, Wei L, Kose K, et al. Real-time video mosaicking to guide handheld in vivo microscopy. *Journal of Biophotonics* 2020.
107. Vercauteren T, Perchant A, Malandain G, Pennec X, Ayache N. Robust mosaicing with correction of motion distortions and tissue deformations for in vivo fibered microscopy. *Medical image analysis* 2006; **10**(5): 673-92.
108. Hu M, Penney G, Rueckert D, et al. A robust mosaicing method with super-resolution for optical medical images. International Workshop on Medical Imaging and Virtual Reality; 2010: Springer; 2010. p. 373-82.
109. Ravi D, Szczotka AB, Shakir DI, Pereira SP, Vercauteren T. Effective deep learning training for single-image super-resolution in endomicroscopy exploiting video-registration-based reconstruction. *International journal of computer assisted radiology and surgery* 2018; **13**(6): 917-24.