Doctoral Dissertations   Student Theses and Dissertations

Spring 2018

Deep learning and localized features fusion for medical image classification

Haidar A. Almubarak

Recommended Citation

DEEP LEARNING AND LOCALIZED FEATURES FUSION FOR MEDICAL IMAGE

CLASSIFICATION

by

HAIDAR ALI ALMUBARAK

A DISSERTATION

Presented to the Graduate Faculty of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

In Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

in

COMPUTER ENGINEERING

2018

Approved by

R. Joe Stanley, Advisor
Randy H. Moss
William V. Stoecker
Donald C. Wunsch II
Bijaya Shrestha
V.A. Samaranayake

PUBLICATION DISSERTATION OPTION

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

Paper I: Pages 5-33 have been published in Information Journal.

Paper II: Pages 34-56 have been submitted to IJHSI Journal.

Paper III: Pages 57-86 have been prepared for submission to IJHSI Journal.

ABSTRACT

Local image features play an important role in many classification tasks as translation and rotation do not severely deteriorate the classification process. They have been commonly used for medical image analysis. In medical applications, it is important to get accurate diagnosis/aid results in the fastest time possible.

This dissertation tries to tackle these problems, first by developing a localized feature-based classification system for medical images and using these features and to give a classification for the entire image, and second, by improving the computational complexity of feature analysis to make it viable as a diagnostic aid system in practical clinical situations.

For local feature development, a new approach based on combining the rising deep learning paradigm with the use of handcrafted features is developed to classify cervical tissue histology images into different cervical intra-epithelial neoplasia classes. Using deep learning combined with handcrafted features improved the accuracy by 8.4% achieving 80.72% exact class classification accuracy compared to 72.29% when using the benchmark feature-based classification method.

ACKNOWLEDGMENTS

I would like to thank my advisor Prof. R. J. Stanley for granting me the opportunity to work under his guidance. His tremendous support and flexibility allowed me to explore and test various ideas in my research and study in the last three years. I would also like to thank my committee members Dr. Moss, Dr. Stoecker, Dr. Wunsch, Dr. Shrestha and Dr. Samaranayake for their help and support and taking time out of their busy schedules and agreeing to serve in my committee. Finally, and most importantly, I'm indebted to my family for their exceptional support. I would like to thank my wife Maryam and my daughters Fawatim and Zainab for putting up with me and waiting for me all these years to finish my degree. I would also like to thank my parents and my wife's parents for their support and helping me taking care of my kids.

TABLE OF CONTENTS

SECTION

LIST OF ILLUSTRATIONS

LIST OF TABLES

SECTION

1. INTRODUCTION

Local image features play an important role in many classification tasks [1, 2]. Translation and rotation do not severely deteriorate the classification process. Local features have been used in several medical image analysis tasks [3, 4, 5] such as improving medical image retrieval significantly [6]. Medical images such as X-ray and CT-scans usually contain local areas of interest used by medical professionals to diagnose and classify different diseases. Within these regions of interest (ROIs), there are some local features that are prominent in different classes of images such as color, size, and shapes in different areas of the image

Analyzing local features from different parts of an image and fusing the analysis result can improve the classification accuracy [7]. The data fusion can either be from different modalities or the same modality but from different regions of the image. Fusing data from different modalities can improve the signal-to-noise ratio of different sensor measurements [8] and can improve detection and classification [9].

One of the main goals of medical image analysis is to develop clinical decision support system, which was shown to improve practitioner performance [10]. One of the important aspects of a clinical decision support system, that is often neglected, is the running time. Medical experts are often overwhelmed by documentation needs and need a fast and accurate overview of each case [11].

For a long time, the classification of medical images involved manually extracting features from the images such as statistical pixel-level features, shape features, texture features and relational features [12]. In recent years deep learning has gained momentum in many image classification tasks because features can be learned from data automatically

[13] and medical image analysis application is no exception. Deep learning techniques have been used in several applications involving medical images, such as pancreas segmentation [14], cytoplasm and nuclei segmentation [15], skin lesion classification [16, 17], mitosis detection in breast cancer images [18], and many more. While deep learning proved to be very helpful in these tasks, manually crafted features still have their place in image analysis and classification. One part of this dissertation demonstrates this by combining handcrafted features and deep learning to achieve higher image classification accuracy than using each paradigm alone.

1.1. PROBLEM STATEMENT

This dissertation explores unique contributions in medical image analysis for localized feature development for localized and image-based classification and for addressing the computational complexity of feature analysis to facilitate computer-assisted diagnostic aids in practical clinical situations. For this work, two medical image classification tasks have been examined. The first was concerned with classifying dermoscopy images into benign or malignant; the color characteristics differences between benign and malignant dermoscopy images are very important in distinguishing them. One issue faced when using color information is that some color ranges are present in both classes. To tackle this problem, the fuzzy-based approach was used and applied to a local cluster of pixels in the dermoscopy images to determine the image class. The second classification task was to classify histology images from cervical tissue for a pre-malignant condition known as cervical intraepithelial neoplasia (CIN) which are seen as atypical cells in the epithelium region of the tissue. A Papanicolaou (Pap) test is used for this type of screening. Definitive evaluation requires histology slides of biopsied cervical tissue, and the interpretation of these slides is done by an expert pathologist [19]. A cervical biopsy is classified as normal (no CIN) or as CIN1 (mild dysplasia), CIN2 (moderate dysplasia), or CIN3 (severe dysplasia) by identifying the atypical cells in the epithelium by the visual inspection of histology slides [20, 21, 22]. In

previous studies [21, 22, 23, 24] it was shown that an epithelium with atypical cells could express different CIN grades in different vertical segments of the epithelium [25]. One way to improve the classification accuracy is to analyze vertical segments of the epithelium independently, and then fuse information from each of these segments to obtain the CIN grade of the whole epithelium region.

1.2. SUMMARY OF CONTRIBUTION

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

1.2.1. Fuzzy Color Clustering for Melanoma Diagnosis in Dermoscopy Images. The research presented is for skin lesion classification using fuzzy logic to fuse/cluster the localized benign and melanoma color information for skin lesion feature development. Color is a very important feature for discriminating skin lesions [26]. In this work, color features were used as the basis for dermoscopic image classification. Clustering is defined as partitioning data objects into a number of clusters or groups based on how similar they are [27]. Proximity measures are used to determine if two objects are similar or not, and the proximity function uses objects' features in the calculation. In this paper, a local cluster of pixels was used in a fuzzy-based algorithm to classify the dermoscopy images into benign or malignant. The two features used to put a pixel in a cluster are: the distance between the the pixel and the cluster centroid, and the fuzzy membership value of the pixel. The color features can appear strongly in one class and weakly in the other class, and hence the fuzzy-based algorithm was used. Fuzzy-based algorithms were shown to improve accuracy in other medical image-related tasks such as segmentation [28] and content-based image retrieval [29]. It was also used for cancer tissue clustering [30].

1.2.2. Algorithm Enhancements for Improvement of Localized Classification of Uterine Cervical Cancer Digital Histology Images. In this research an improvement to the computational efficiency of feature extraction and image analysis techniques devel-

oped in [31] was examined to promote epithelium CIN classification in an expert system development with application in a clinical setting. A graphic user interface was developed to be used by clinicians, because the original algorithm was very slow and not suitable for a clinical setting, the algorithm was examined thoroughly , and the bottlenecks were identified. Several changes were made and tested, resulting in 18 experiments. The modification improved the time significantly while maintaining the classification accuracy.

1.2.3. Classification of Uterine Cervical Cancer Digital Histology Images using A Hybrid Deep Learning and Handcrafted Features Approach. Deep learning in the form of a deep convolutional network proved to be great in many application areas related to images, such as image classification [32], face recognition [33], breast cancer mitosis detection [18] and many more. An important factor in rise of deep learning adoption is that it does not need engineered features for classification. Despite this important element of deep learning, handcrafted features are still useful and important. In this research, a new data fusion paradigm was developed based on deep learning and handcrafted features techniques for localized and entire image classification for cervical intra-epithelial neoplasia in digitized histology images. Using a deep learning approach alone gave a classification accuracy similar to the handcrafted features approach, but combining the two approaches yielded better classification results.

PAPER

I. FUZZY COLOR CLUSTERING FOR MELANOMA DIAGNOSIS IN DERMOSCOPY IMAGES

Haidar A. Almubarak[1], R. Joe Stanley[1], William V. Stoecker[2], and Randy H. Moss[1]

[1]Department of Electrical & Computer Engineering

Missouri University of Science and Technology Rolla, Missouri 65409–0050

Tel: 573–341–6896

[2]Stoecker and Associates, Rolla, MO

Email: h.almubarak@ieee.org, {stanleyj, rhm, wvs}@mst.edu

ABSTRACT

A fuzzy logic-based color histogram analysis technique is presented for discriminating benign skin lesions from malignant melanomas in dermoscopy images. The approach extends previous research for utilizing a fuzzy set for skin lesion color for a specified class of skin lesions, using alpha-cut and support set cardinality for quantifying a fuzzy ratio skin lesion color feature. Skin lesion discrimination results are reported for the fuzzy clustering ratio over different regions of the lesion over a data set of 517 dermoscopy images consisting of 175 invasive melanomas and 342 benign lesions. Experimental results show that the fuzzy clustering ratio applied over an eight-connected neighborhood on the outer 25% of the skin lesion with an alpha-cut of 0.08 can recognize 92.6% of melanomas with approximately 13.5% false positive lesions. These results show the critical importance of colors in the lesion periphery. Our fuzzy logic-based description of lesion colors offers relevance to clinical descriptions of malignant melanoma.

Keywords: Image Processing, Dermatology, Color, Malignant Melanoma, Histogram, Fuzzy Logic

1. INTRODUCTION

Diagnosis of early-stage melanomas is challenging, especially for melanomas which have an atypical presentation [1, 2, 3, 4]. Dermoscopy is an imaging modality that uses a simple hand-held device that eliminates surface glare and magnifies structures invisible to the naked eye. Although dermoscopy significantly improves diagnostic accuracy compared to naked-eye examination alone [5, 6], average dermoscopic melanoma sensitivity for dermatologists and others seeing dermatology patients in three recent reader studies was only 71 - 85% and specificity 54 - 71% [7, 8, 9]. Non-specialist sensitivity and specificity is even lower; when compared to dermatologists, non-specialists were found to have a sensitivity for melanoma of 54% and a specificity of 73% [10].

The relatively low diagnostic accuracy for discrimination of malignant melanoma demonstrates the uncertainty involved in skin lesion analysis. This low diagnostic accuracy is an important medical problem for society, since the incidence of cutaneous melanoma, 161,790 cases estimated in the US in 2017, vs. 140,860 cases for 2016, has risen every year since 1979 [11, 12]. In more recent advances, techniques such as deep learning have shown potential using pixel blocks from lesion images with disease labels for significantly improved skin lesion discrimination, achieving an area under the diagnostic ROC curve of approximately 0.91, exceeding the performance of dermatologists [8]. Even with the improvements in computer-assisted techniques for skin lesion classification, learning and application of specific features for skin lesion discrimination is important for clinicians and clinician training. Therefore, the investigation of computer-assisted techniques for the determination and analysis of specific skin lesion features is useful in the dermatology community. In this research, we focus on automatic detection of melanoma colors.

One of the most important features for detecting malignant melanoma is the skin lesion color. Some studies have examined the influence of shape and color for skin lesion discrimination and determined that shape is an important contributor for successful lesion assessment [13]. Contributors to this assessment may include melanomas that tend to have more colors present, often with different shades of tan, brown, brown or black and occasional patches of red, white or blue. Moreover, melanoma colors are present in chaotic patterns, classically described as "variegated" [14]. Many analytic color descriptors have been applied to melanoma detection, most often utilizing the red, green and blue (RGB) color space. These descriptors are among the most significant of all analytic descriptors of melanoma. Ferris et al., found that the top three statistical features for melanoma were all color feature characterizing color histograms and color asymmetry [7]. Analysis of various analytic and clinical features by Rubegni et al., showed that the single most significant feature was red asymmetry [15]. This research group used variables that capture the number and types of color elements within a lesion, to represent the chaotic pattern of color present within a lesion, related to the "color islands" previously described by this group [16]. Other color descriptors used to detect melanoma include variation of hues [17], analytical color techniques for detecting color variegation [18], RGB color channel statistical parameters [19, 20, 21], spherical color coordinates and (L, a*, b*) color coordinate features [22], percentage of the skin lesion containing absolute shades of reddish, bluish, grayish and blackish areas and the number of those color shades present within the skin lesion [23]. Color quantization for the different color shades examined in [24] was performed using the median cut color quantization algorithm [24]. Using a training set of lesions to populate a cumulative relative color histogram based on clinical melanoma and benign images [25, 26] and dermoscopy images [27] has been explored for determining melanoma and benign colors based on probabilistic and histogram region growing techniques for finding the percentage of pixels within the skin lesion that are melanoma colored [25] and the clustering of melanoma colored pixels within the lesion [26, 27]. Melanoma discrimination also has

been investigated using fuzzy logic techniques to quantify the degree of association of lesion colors with benign lesions based on determining a membership function for benign color derived from a training set of benign lesions [28]. Color features sometimes are combined with other features such as texture, but when comparing texture and color alone the color features outperform the texture features [29].

Using color to characterize melanoma is faced with two major difficulties. First, variations in lighting and slide processing make it difficult to characterize melanoma. The use of relative color [27, 30, 31], where the average background skin color is subtracted from each lesion pixel, has been used to help compensate for color variability in the imaging process. The use of the relative color technique tries to remove variability of color due to different skin types as well as to lighting and image processing techniques [30]. Second, different melanomas have different color characteristics across skin lesions which makes it difficult to reliably use melanoma color features as distinct differentiators of melanoma. These difficulties highlight the need for color analysis techniques that do not explicitly identify colors representative of melanoma (or benign) lesions as the foundation for color-based skin lesion diagnosis. Fuzzy logic techniques can represent color with these inherent difficulties through quantization of degrees of association as melanoma or benign colors for skin lesion discrimination. This study extends the data-driven approaches from [25, 26] using a training set of images for determining relative color degree of association or membership as a benign color for finding the ratio of pixels within a skin lesion with a specified level of membership as a benign color to all pixels within the lesion that have a non-zero membership as a benign color [28]. This study introduces fuzzy logic techniques for relative color representation that incorporates spatial context for the clustering of pixels within specified regions of the skin lesion for color feature analysis. The goal of this study is to present advances in color features for melanoma and benign lesion discrimination that can be useful for clinicians in the diagnostic process.

The remaining sections of the paper include: (1) an overview of dermoscopy image acquisition; (2) the fuzzy logic technique for color histogram analysis and color feature determination; (3) an overview of the lesion region analysis; (4) the automated approach for melanoma detection; (5) experiments performed; (6) results and discussion and (7) conclusions.

2. METHODOLOGY

2.1. Dermoscopy Image Data. Imaging of skin lesions is anticipated to have an important role as a diagnostic aid for malignant melanoma. In this research, the red, green and blue (RGB) color space of dermoscopy images was used to analyze skin lesions. In the RGB space each dimension (R, G and B) typically uses 8 bits per pixel, providing 256 possible values for each dimension. Thus, there are a total of 256^3 or 16,777,216 possible colors in this color space.

This research uses dermoscopy skin lesion images with a typical size of 1024 x 768 pixels; all are digitized in RGB color space. The data set was selected at random from 2000 dermoscopy images from Skin and Cancer Associates (Plantation, FL) and the Dermatology Associates of Tallahassee, FL, in addition to images acquired from the EDRA Interactive Atlas of Dermoscopy [32]. The images from the EDRA Interactive Atlas of Dermoscopy were acquired from clinics in Europe using Heine Dermaphot equipment with various cameras. Both sets of images were obtained using non-polarized contact dermoscopy techniques. The study set includes 175 malignant melanoma images and 342 dysplastic nevus (benign) images. The melanomas are all invasive malignant melanomas, with no melanomas in situ. Thus the set of images in this study samples a heterogeneous population from a wide geographic area obtained with different techniques. Additionally, dysplastic nevi in the image set show some degree of "atypia", i.e., they have some degree of the color chaos/variegation found in melanomas. Thus the images chosen comprise a challenging diagnostic set of lesions. Figure 1 shows image examples from the image data

set of a malignant melanoma (a) and a benign (dysplastic nevus) lesion (b). These images demonstrate the degree of the degree of difficulty of diagnostic assessment in this group of lesions.

<div align="center">(a)   (b)</div>
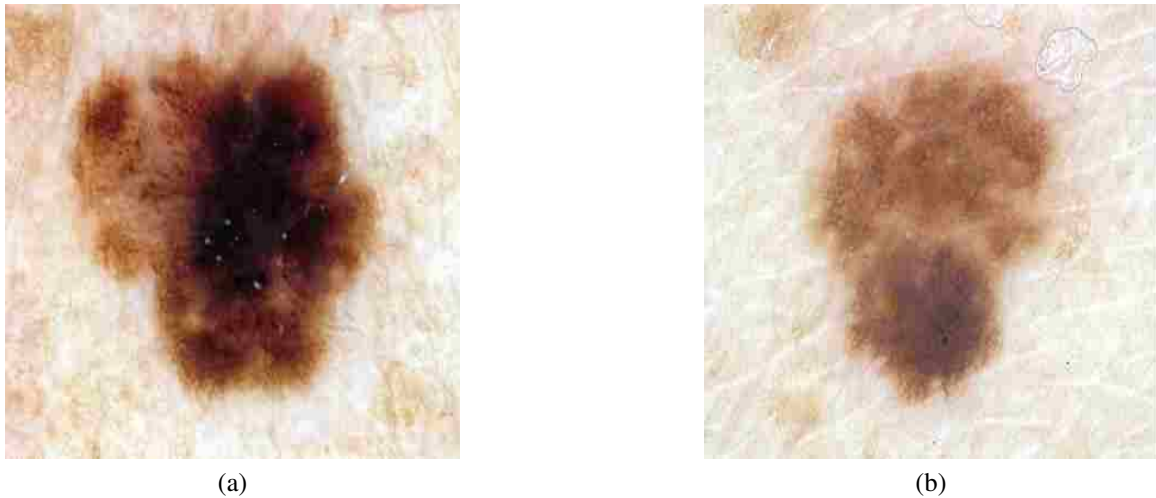
Figure 1. Example of Malignant Melanoma (a) and Benign Lesion (b)

2.2. Diagnostic Assessment. Table 1 below shows the diagnostic interpretations of skin lesions for evaluating the automated skin lesion classification results that were applied in this investigation. A true positive (TP) means diagnosing a melanoma as a melanoma. A false negative (FN) means diagnosing a melanoma as a benign skin lesion. A true negative (TN) means diagnosing a benign skin lesion as benign. A false positive (FP) means diagnosing a benign skin lesion as a melanoma. The implication of a false negative is delaying therapy for a melanoma, with worse consequences for the patient, while a false positive results in an unnecessary biopsy.

2.3. Skin Lesion Border Determination. In this research, skin lesion borders are found using the following procedure. A program called xborder developed by McLean [33] is used to choose points along the border manually. After choosing the points, a closed curve is formed using the points; the curve minimizes the spline curve function's second derivative. A dermatologist, working with a group of students, verified all borders drawn; modification to the borders were done according to the dermatologist's recommendation. The process

Table 1. Skin Lesion Diagnosis Categorization Based on the Type of Skin Lesion

	Computer-Assisted Skin Lesion Diagnosis	
Actual Skin Lesion Diagnosis	Benign	Melanoma
Benign	True negative (tn)	False positive (fp)
Melanoma	False negative (fn)	True positive (tp)

therefore sought to find as accurate borders as possible without a known gold standard for border accuracy. This approach has been used in several skin lesion feature analysis studies [22, 25, 26, 27, 28, 33, 34, 35]. For an M row by N column RGB image, I, each pixel in position (x,y) can be represented by a 3-tuple $I(x, y) = (R(x, y), G(x, y), B(x, y))$, where $1 \leqslant x \leqslant N$ and $1 \leqslant y \leqslant M$.

2.4. Relative Color. Differences in lighting conditions, cameras, film types, and skin colors should be taken into consideration when building a classification system of skin lesions; the system needs to be relatively tolerant to these variations. This is particularly true for our data set which came from three sources of lesions using digitized 35 mm photographic slides; we had no control over cameras used to capture original photographed skin lesion data. In order to make relative tolerance to these variations possible, a relative color method was used. The relative color of a skin lesion pixel is defined as the difference between that particular pixel value and the average value of pixels in the skin surrounding the lesion [28]. Let ROI denote the skin lesion region in the image and be referenced as $ROI = \{(x, y)|(x, y) \in I$ and $(x, y) \in$ `area inside the closed skin lesion boundary`$\}$. Then, the relative color for all skin lesion pixels is given as

$ROI_{rel}(x, y) = (R_{rel}(x, y), G_{rel}(x, y), B_{rel}(x, y)) = (R(x, y) - R_{skin}, G(x, y) - G_{skin}, B(x, y) - B_{skin}|(x, y) \in ROI)$, where $-255 \leq R_{rel}(x, y), G_{rel}(x, y), B_{rel}(x, y) \leq 255$. R_{skin}, G_{skin} and B_{skin} are the average red, green and blue plane pixel values computed from the surrounding skin. Section 2.5 presents the procedure for finding the surrounding skin. Relative color provides many advantages when used for analyzing skin lesions. First, differences resulting

from variation in ambient light and their respected digitization errors are reduced when relative color is used. Second, errors in digitizing images from different film types or different film processing techniques can be minimized by relative color. Third, the mammalian visual system operates in a manner similar to the relative color technique. Finally, skin color varies among persons and relative color tries to minimize this variation and produce a skin-tone indifferent system

2.5. Surrounding Skin Color Determination. Several empirical relationships were utilized to determine non-skin color, pixels in deep shadow, direct flash reflection and exogenous marks are eliminated. An existing dermatology image database was used under the supervision and guidance of a dermatologist to derive the skin pixel finder used in this research. The skin pixel finder has been applied to skin lesion analysis in other research [28, 34]. To approximate the surrounding skin color, the distance function is used twice. First, a 10-pixel wide area surrounding the lesion is removed from consideration, to allow for possible inaccuracy in border determination. Then the distance function is used to determine an annular region just outside the 10-pixel-wide region; the area of this outer annular region is 20% of the skin lesion size [28, 34], which is the area used for surrounding skin calculation. If the lesion extends close to the image edge, the remaining non-lesion pixels within this annular region are used rather than the standard 20% of lesion size. The pixels determined by this technique are used to determine the average surrounding skin color (R_{skin}, G_{skin} and B_{skin}) pixel values.

2.6. Color Histogram Bin Determination. The relative color histogram for the set of benign lesions is used to develop the fuzzy logic techniques used in this research. The pixel values for R_{rel}, G_{rel} and B_{rel} in the relative color space,defined by subtracting the R, G and B values from the surrounding skin, can take values in the range [-255,255]. Pixels from the entire benign lesion training set thus have 511 possible values on each of the three color axes, resulting in a three-dimensional histogram for counting pixel values; this histogram has 511^3 possible values. To increase statistical significance for benign-malignant color

separation and reduce computational complexity, these values are quantized by dividing each original relative color value by 4, to create the final relative color histogram with dimensions 128 x 128 x 128. Each final quantized histogram value represents a cube of dimension 4 X 4 X 4 containing up to 64 original lesion color values. Thus each point in the final relative color histogram is a "bin" that counts pixels with up to 64 possible original color values.

Formally, let C_{bin} denote the set of relative color bins, and let $C_{bin}(x, y)$ denote the relative color bin in which the pixel (x, y) of the skin lesion falls. Since each of the original axes has 511 possible values, quantization by dividing the axis by 4 will result in bins on the edges of the cube that are 4 x 4 x 3 and one bin in the corner of the cube that is 3 x 3 x 3. Upon examining the training set relative color histograms, the bin that contains the relative color -255 (before quantization) will not likely occur in dermoscopy images in the three planes R, G, and B; hence that cube is assigned the smaller dimension (3^3).

2.7. Color Histogram Analysis Technique. The color histogram analysis consists of three steps: Fuzzy set description, Color feature determination, and Threshold determination. The following sections describe each step in detail.

2.7.1. Fuzzy set description for trapezoidal membership function. The relative skin lesion color was represented using a fuzzy logic-based approach, which was initially developed and reported in [28]. Let **B** represent a fuzzy set [36] having a trapezoidal membership function for relative skin lesion color, for the specified skin lesion class. This study utilizes the same trapezoidal membership function from [25], which is derived using a data-driven process that assigns membership values to skin lesion relative colors based on frequency of occurrence in the training set of skin lesions. The colors within the color histogram bins defined in the Section 2.6 were assigned membership values using the following process: The training set of images for the specified class is used to populate the three-dimensional relative color histogram bins in batch mode; each bin will contain the sum of pixels with relative color mapping to that bin from all skin lesion regions in all the

training images. A secondary histogram is defined as a histogram of the three-dimensional relative color histogram. The secondary histogram is a function of bins in the relative color histogram; it will indicate the number of bins containing x pixels in the relative color histogram. The benign skin lesion from the training data is used to determine the fuzzy set **B**. The secondary histogram of the relative colors is used to assign the membership values. The membership function $\mu_B(x)$ denoting the fuzzy set [36] is given as

$$\mu_B(x) = \begin{cases} \frac{x}{F} & \text{for } 0 \leq x < F \\ 1 & \text{for } x \geq F \end{cases} \tag{1}$$

where x is the bin frequency in the secondary histogram and F is the empirically determined bin frequency count for full membership in the benign fuzzy set **B**. F is set as the frequency count given that 5% of the secondary histogram of the benign skin lesions have frequency F or greater. Increased bin frequency (up to F) will be reflected as increased membership value in the specified class of skin lesion. Figure 2 shows a representative secondary histogram with the frequency count F labeled in (a) and the trapezoidal membership function generated in (b). The horizontal axis provides the frequency of occurrence (x) that a bin is populated over all benign images of the training set. The vertical axis in Figure 2a gives the number of bins with x "hits" per bin over the training set of benign images.

The trapezoidal shape compensates for the disparities in the 4 x 4 x 4 relative color histogram bins populations from the training set of images where relative color histogram bins with lower hits that are spatially near relative color histogram bins with a higher number of hits are designated with degrees of association (i.e. membership) as benign colors. These lower hit relative color histogram bins would be ignored in a crisp labeling, where a binary decision is made to include or exclude a relative color bin as a benign color.

2.7.2. Color feature determination. The level of confidence of being in a skin lesion class **B** is given by the fuzzy clustering ratio feature. The fuzzy clustering ratio feature is computed as follows. Let **M** denote the set of relative color values that map into
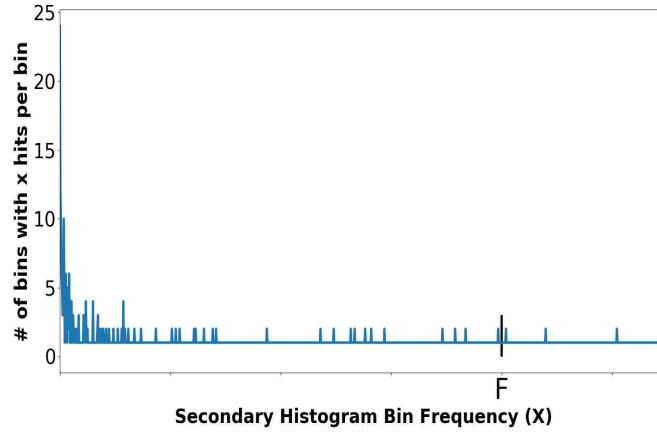
(a)
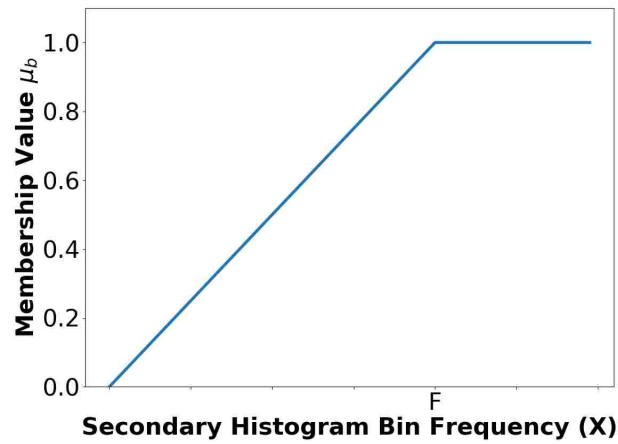
(b)

Figure 2. A Relative Secondary Histogram (a) and Its Corresponding Trapezoidal Membership Function for Fuzzy Set B (b). The Frequency Count F is Labeled on the Secondary Histogram and Membership Function Plot.

relative color histogram bins labeled with

$$\mu_B(C_{Bin}(x, y)) \geq \alpha \tag{2}$$

for a specified α on **B** and (x, y) contained in the skin lesion. **M** is given by

$$M = \left\{ ROI_{rel(x,y)} : \mu_B(C_{bin}(x, y)) \geq \alpha \text{ for}(x, y) \in ROI \right\} \tag{3}$$

Let P denote the set of relative color values that map into relative color histogram bins labeled with

$$\mu_B(C_{bin}(x, y)) > 0 \tag{4}$$

and (x, y) contained in the skin lesion such that

$$P = \left\{ ROI_{rel(x,y)} : \mu_B(C_{bin}(x, y)) > 0 \text{ for} (x, y) \in ROI \right\} \tag{5}$$

Let **L** denote the set of pixel locations within the skin lesion region of interest with relative colors $ROI_{rel(x,y)}$ that map into **M**, formally

$$L = \left\{ (x, y) | ROI_{rel(x,y)} \in M \right\} \tag{6}$$

Let $N_{((x,y),r)}$ denote the number of 8-connected neighbors within the square neighborhood of radius r for $(x, y) \in ROI$ and $N_{(M(x,y),r)}$ denote the number of neighbors connected within the square neighborhood of radius r that are contained in the lesion region of interest for pixel $(x, y) \in L$, such that $r = 0$ refers to the pixel (x, y) itself, $r = 1$ refers to the eight-connected neighborhood of (x, y) and so on. By Equation (6), the neighbors of radius r for a point $(x, y) \in L$ are excluded from calculating $N_{(x,y)}$ and $N_{(M(x,y),r)}$ if the neighbors lie outside of the lesion ROI. Then,

$$S(\alpha, r) = \sum_{(x,y) \in L} N_{(M(x,y),r)} \tag{7}$$

represents the total number (running sum) of neighbors for all pixels within the skin lesion with relative color values that map as colors of the specified class at the specified alpha-cut. The cumulative total number of eight-connected neighbors for all $(x, y) \in L$ is denoted as

$$V(r) = \sum_{(x,y) \in P} N_{((x,y),r)} \tag{8}$$

V includes all neighbors of pixels within the skin lesion that have a non-zero membership value in **B** for the specified lesion class. The fuzzy clustering ratio for a skin lesion is given as

$$R(\alpha, r) = \frac{S(\alpha, r)}{V(R)} \tag{9}$$

Note that $R(\alpha, r)$ is computed based on the pixels within the skin lesion. The surrounding skin pixels are only used for computing the average surrounding skin color for determining the relative color of all pixels within the skin lesion. Thus, for $\alpha = 0$, $R = 1$, provided that at least one pixel within the skin lesion has a non-zero membership value in **B** for a given radius r. If **B** represents the fuzzy set for benign skin lesion relative color as determined from the benign training set of images, then $R(\alpha, r)$ represents the degree to which the clustering of benign colors within square radius r of the skin lesion are the colors perceived to be associated with benign skin lesions. After determining **B**, the next step is to compute $R(\alpha, r)$ for all benign and melanoma skin lesions from the training data for a specified value of α. A threshold T is automatically determined from the ratios $R(\alpha, r)$ calculated from the training data. The procedure for finding T is presented in the following Section 2.7.3. Skin lesions are categorized as either benign or melanomas for the data set used in this research. A given skin lesion is classified as benign if $R(\alpha) > T$. Otherwise, the skin lesion is labeled as a melanoma. A similar process is used to determine **B**, $R(\alpha)$, and T using melanoma skin lesion images as the training set. However, a given skin lesion would be classified as benign if $R(\alpha, r) \leq T$.

2.7.3. Threshold determination procedure. The ratios $R(\alpha)$ computed from the training data were sorted to facilitate automated threshold (T) selection. The approach used for automatically selecting T for a particular is based on computing the true positive and true negative rates for the training data. The procedure for choosing the optimal T involves iterating T through the sorted ascending ratios $R(\alpha, r)$ from [0,1] in increments of 0.001. For each T, tp and tn are determined from the training images. T is chosen as the greatest

threshold for which tn ≤ tp. The final melanoma and benign lesion classification results are obtained for the training and the test data using the final threshold T. The procedure is repeated for specified values.

2.8. Lesion Region Analysis. After determining $\mathbf{B}$ and μ_B from the training set of images for the specified class, $R(\alpha, r)$ is computed over different skin lesion regions, referenced as boundary area percentages, for skin lesion discrimination. The boundary area percentage is defined as the uniform skin lesion region closest to the boundary that contains a specified percentage of the lesion area. The boundary area region is found using the method developed in [37] with summary in pseudocode:

1. start with the lesion border

2. compute the distance transform using 8-connected distance to get pixel distance from the lesion boundary inside of the skin lesion

3. starting with a distance of 0 (lesion boundary), determine the lesion boundary region with distance less than or equal to the current distance

4. compute the area of the resulting lesion boundary region

5. retain the lesion boundary region if its area is greater than the boundary area percentage

6. otherwise, increment the distance by 1 and repeat Steps (2) - (6)

Steps (5) and (6) show that the process continues until the area percentage criterion has been satisfied. Figure 3 shows examples of boundary area percentages of 10%, 25% and 75% of the lesion area; only the areas shown in white are used for feature calculations; the remaining lesion interior is excluded from feature calculations.
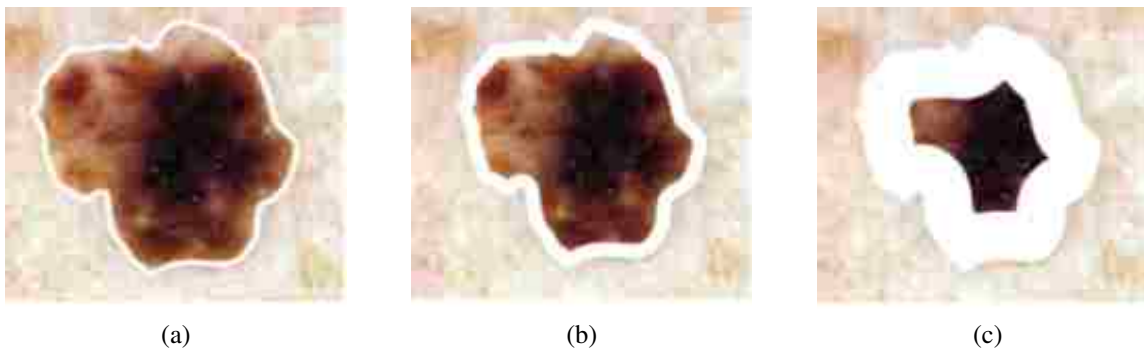
Figure 3. Boundary Area Percentage Examples Using 10% (a), 25% (b) and 75% (c) of the Lesion Area for Analysis (White Region). This is the Same Lesion as in Figure 1

3. EXPERIMENTS PERFORMED

For this research, 175 melanoma and 342 benign dermoscopy images are analyzed. The fuzzy logic-based analysis technique is evaluated on 25 randomly chosen training/test sets with a 90/10 ratio. The relative color fuzzy set **B** with trapezoidal membership function is determined from the benign training images using the entire skin lesion. Variables are assigned as follows: $\alpha = 0.1, 0.2, 0.4, 0.6, 0.8$ and 1, $r = 0$, and 1 and boundary area percentage 10%, 25%, 75%, and 100%; classification of test lesions as melanoma or benign is based upon the optimal threshold **T**, Section 2.7.3. The entire process is repeated, this time using the melanoma training images to determine **B**.

4. RESULTS AND DISCUSSION

Figure 4 presents the average test results over 25 trials for $r = 0$ boundary area percentages of 10%, 25%, 75%, 100% in (a) - (d), respectively. Figure 5 shows the average test results over 25 trials for $r = 1$ and $\alpha = 0.1, 0.2, 0.4, 0.6, 0.8, 1.0$ for boundary area percentages of 10%, 25%, 75%, 100% in (a) -(d), respectively. In Figures 4 (a) - (d) and 5 (a) - (d), the horizontal axis provides the alpha-cut value, and the vertical axis gives the correct percentage discrimination rate. For each alpha-cut, the following labeling scheme

is used: (1) the average correct melanoma rate found based on computing $R(\alpha, r)$ using the benign training images for determining the fuzzy set **B** is denoted as FCN-Melanoma; (2) the average correct melanoma rate found based on computing $R(\alpha, r)$ using the melanoma training images for determining the fuzzy set **B** is denoted as FCM-Melanoma; (3) the average correct benign rate found based on computing $R(\alpha, r)$ using the benign training images for determining the fuzzy set **B** is denoted as FCN-Benign; and (4) the average correct benign rate found based on computing $R(\alpha, r)$ using the melanoma training images for determining the fuzzy set **B** is denoted as FCM-Benign. Note that other boundary area percentage areas, including 50%, were tested, but the results were not improved over the tables presented in this study.

The experimental results shown in Figures 4 and 5 highlight several observations. First, from Figure 4 the average correct melanoma and benign rates are higher when computing $R(\alpha, r = 0)$ based on determining **B** using the benign training set of images than the rates based on the melanoma training set of images for all boundary area percentage cases examined. In particular, the disparity between the melanoma correct rates for the benign and melanoma training approaches, respectively, for determining $R(\alpha, r = 0)$ increases with the boundary area percentage examined, with the greatest disparity observed for the 100% boundary area percentage case (entire lesion). A similar trend is observed between the benign correct rates for the benign and melanoma approaches, respectively. Using the benign training images to determine **B** for computing $R(\alpha, r = 0)$ may yield better discrimination capability because relative colors may be more consistent in benign lesions than in melanoma lesions.

Second, from Figures 4 and 5 for $r = 0$ and $r = 1$, the test true positive rate (FCN-Melanoma) exceeds the true negative rate (FCN-Benign) for all boundary percentage area cases using the benign training images to determine the fuzzy set B.
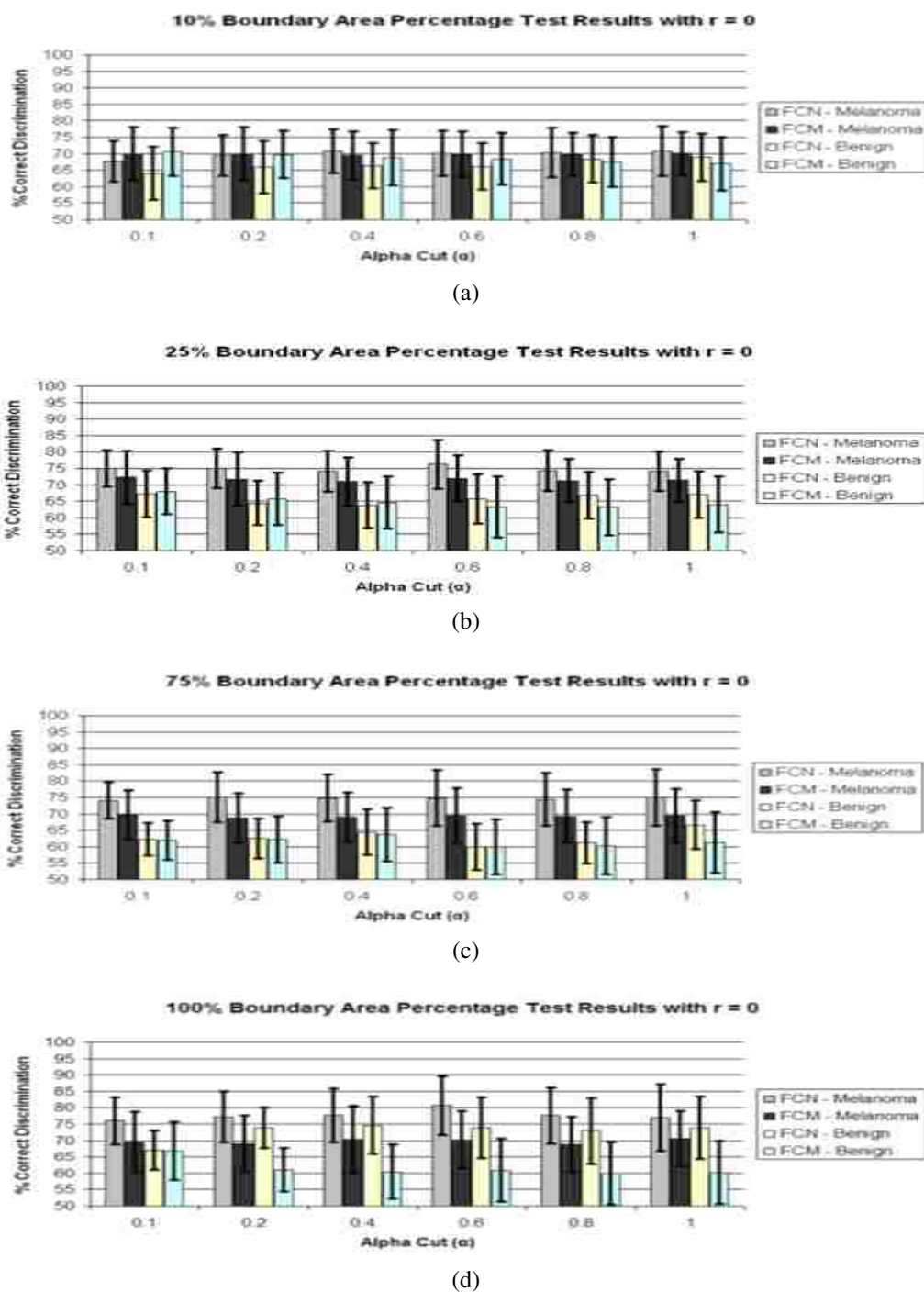
Figure 4. Average and Standard Deviation Test Results Over 25 Randomly Generated Training/Test Sets for Different Boundary Area Percentages for r = 0. FCN and FCM Denote The Fuzzy Ratio Features Computed Using the Benign and Melanoma Training Images to Determine Discrimination Accuracy. (a) Boundary Area Percentage = 10% (b) Boundary Area Percentage = 25% (c) Boundary Area Percentage = 75% (d) Boundary Area Percentage = 100%

10% Boundary Area Percentage Test Results with r = 1

(a)

25% Boundary Area Percentage Test Results with r = 1

(b)

75% Boundary Area Percentage Test Results with r = 1

(c)

100% Boundary Area Percentage Test Results with r = 1
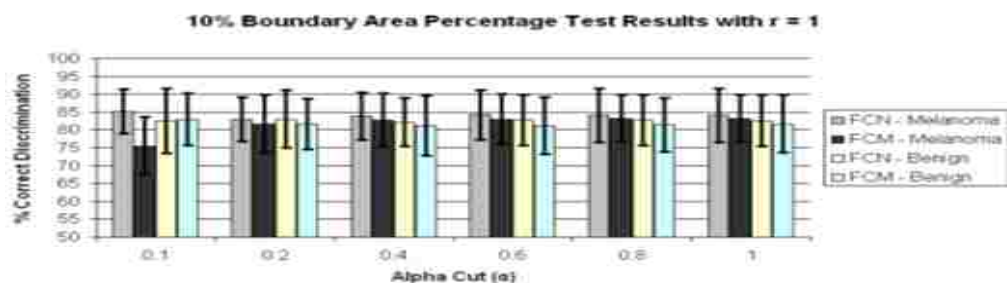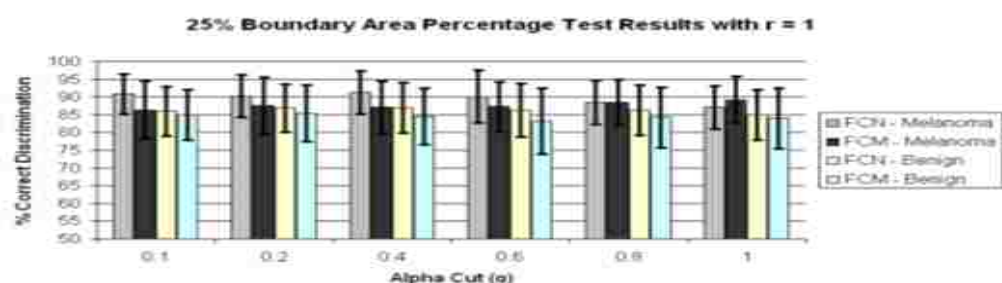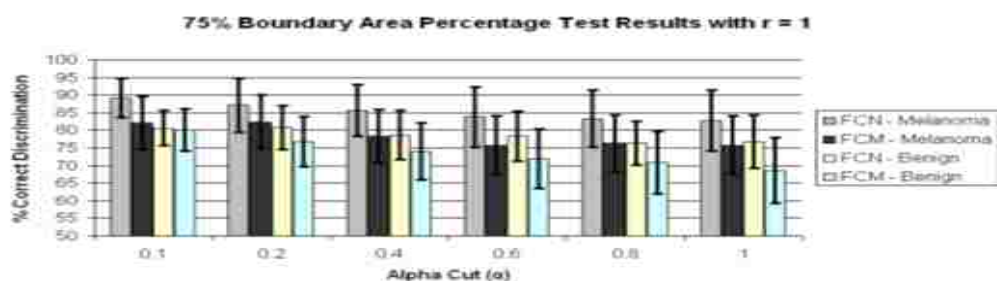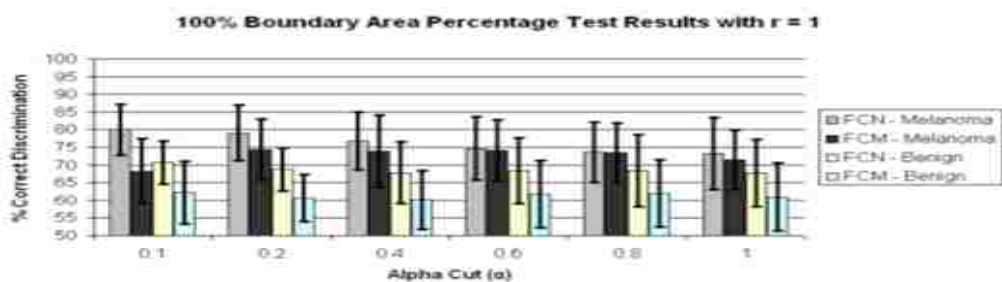
(d)

Figure 5. Average and Standard Deviation Test Results Over 25 Randomly Generated Training/Test Sets for Different Boundary Area Percentages for r=1. FCN and FCM Denote the Fuzzy Ratio features computed using the benign and melanoma training images to determine discrimination accuracy. (a) Boundary area percentage = 10% (b) Boundary area percentage = 25% (c) Boundary area percentage = 75% (d) Boundary area percentage = 100%

For fuzzy ratio $R(\alpha, r)$ calculations, the numerator $S(\alpha, r)$ and the denominator $V(r)$ are dependent on the training set for the specified class of images used to determine $\mathbf{B}$. Every pixel within every training image of the specified class of images has a nonzero membership in $\mathbf{B}$. Accordingly, $V(r)$ is equal to the lesion area for those training images, and each pixel within every training skin lesion is included in $S(\alpha, r)$ for some range of $\alpha > 0$. This is not necessarily true for the non-specified class of training images and the melanoma and benign test images. Because the training fuzzy ratios $R(\alpha, r)$ for the specified class of training images are slightly skewed from the test benign fuzzy ratios $R(\alpha, r)$, the threshold $\mathbf{T}$ determined from the training set of images may not be as reflective of the test fuzzy ratios for the specified class of images as for the non-specified class of images. A similar trend is observed from Figures 4 and 5 for $r = 0$ and $r = 1$ for the true positive (FCM-Melanoma) and true negative (FCM-Benign) rates for all boundary percentage area cases except for the 10% case using the melanoma training images to determine $\mathbf{B}$, where the true positive rate exceeds the true negative rate for $\alpha > 0.4$. This trend could be attributed to the approach used for deriving the fuzzy set $\mathbf{B}$ and the membership function μ_B. There are fewer melanoma images than benign lesion images, and the 10% boundary area cases results in fewer relative color histogram bins being populated, yielding most relative colors as having low membership in $\mathbf{B}$. Thus, melanoma and benign lesions have similar $R(\alpha, r)$ for small values of $\alpha(\alpha < 0.4)$, leading to low melanoma discrimination rates.

Third, for $r = 0$ the highest true positive and true negative rates are achieved for the 100% boundary area case, with the highest melanoma discrimination rate of 80.8% at $\alpha = 0.6$ and the highest benign discrimination rate of 74.6% at $\alpha = 0.4$. The $r = 0$ case corresponds to computing the fuzzy ratio introduced in previous research for clinical images [28]. These results for the dermoscopy images are slightly lower than the true positive and true negative results achieved applying this feature to skin lesion discrimination in clinical images, where the highest true positive results obtained were 91.74% with a corresponding true negative rate of 77.49% at $\alpha = 0.05$ [28]. The dermoscopy image results for $r = 0$ are

likely lower compared to the clinical image results reported in [25] because there are wider color variations in dermoscopy images than in clinical images. The resulting cumulative histograms from the training set of lesions have fewer relative color bins with a high number of hits, so there are a smaller number of relative histogram color bins with high number of hits, so there are a smaller number of relative histogram color bins with high membership. This particularly true for the melanoma lesion images, which tend to have greater color variation than benign lesion images. The $r = 0$ case provides for all pixels within the lesion that satisfy the alpha-cut constraint in the numerator of $R(\alpha, r)$ and provide for any degree of membership in B in the denominator $V(r)$ without taking into account spatial context for feature determination. As a result, there is less variation in $R(\alpha, r = 0)$ for the benign and melanoma images for different values of α. The experiments for computing $R(\alpha, r = 1)$ incorporate only some spatial context for feature calculations which allows for clustering of neighboring pixels with membership values that satisfy the constraint.

Fourth, for $r = 1$ the highest true positive (FCN-Melanoma) and true negative (FCN-Benign) rates are achieved for the 25% boundary area case using the benign images as the training set of images to determine **B**, with the highest true positive rate of 91.29% at $\alpha = 0.4$ and the highest true negative rate of 86.9% at $\alpha = 0.2$ and 0.4. The true negative rates using the benign images as the training set of images are consistently higher at the 25% boundary area case for all alpha-cuts than for the other boundary area percentage cases. The feature $R(\alpha, r)$ for $r = 1$ is computed as the sum of the number of eight-connected neighbors that satisfy the alpha-cut criterion for each pixel within the lesion that satisfies the alpha-cut criterion divided by the sum of eight-connected neighbors that have non-zero membership in B for each pixel within the lesion that has non-zero membership in **B**. This feature is similar to a crisp color clustering ratio investigated in previous research [26, 27]. In order to determine if any improvement in lesion discrimination can be achieved based on refining the alpha-cut criterion, an experiment was performed for the $r = 1$ case using $\alpha = 0.02, 0.05, 0.08$, and 0.15. Figure 6 shows the average and standard deviation

true positive (FCN-Melanoma) and true negative (FCN-Benign) results for $r = 1$ and $\alpha = 0.02, 0.05, 0.08, 0.1, 0.15$, and 0.2 based on using the benign images as the training set of images to determine **B**. Note that the true positive and true negative results for $\alpha = 0.1$ and 0.2 are contained in Figure 5b. From Figure 6 the highest true positive rate of 92.6% is obtained for $\alpha = 0.08$ with a corresponding true negative rate of 86.5%. This combination of $r = 1$ and $\alpha = 0.08$ also yielded the highest overall weighted true positive and true negative correct recognition rate of 88.51%, compared to the combination $r = 0$ and $\alpha = 1$ giving the highest weighted correct recognition rate of 73.73%. These results highlight the significant lesion discrimination capability of the fuzzy clustering ratio, incorporating spatial information using neighbor clustering ($r = 1$) over individual pixel-based color descriptor determination (fuzzy ratio) for enhancing skin lesion discrimination.

The color clustering ratio is based on using both melanoma and benign training sets of images for populating a cumulative relative color histogram, probabilistic labeling of the relative color histogram bins as melanoma or benign colors with melanoma and benign color region growing within the histogram for final melanoma and benign color bin assignments. The clustering ratio feature is computed over the boundary region of interest to determine clustering of melanoma colors. For each melanoma pixel, i.e., for each pixel that maps as a melanoma color, the sum of eight-connected pixels that map as melanoma colors is calculated and added to the cumulative sum for all melanoma pixels in the boundary region. This total is divided by the total of all eight-connected neighboring pixels for all melanoma pixels in the region. Applying the color clustering ratio to skin lesion discrimination in dermoscopy images yielded true positive, true negative, and overall classification rates as high as 87.7%, 74.9%, and 81.3%, respectively, for the 75% boundary area percentage case [27]. The fuzzy logic-based approach improves the skin lesion discrimination capability over the crisp clustering ratio. The fuzzy logic-based approach achieves the true positive highest results in the outer 25% of the lesion, which is consistent with the color clustering ratio results obtained for clinical images [26]. The utility of the fuzzy logic-based approach

is that the model used for feature calculations is data-driven based on the class of interest to determine the fuzzy set **B** and its corresponding membership function μ_B and only requires information about the remaining class to find the threshold **T** for discrimination. In the feature calculation process, no decision is necessary to label colors as melanoma or benign. Rather, the data-driven process provides for the degree of association of different colors as melanoma or benign, depending on the class used to determine **B**, based on the frequency of occurrence in the training set of images. Because relative color is used for mapping RGB colors at each pixel location in the lesion to relative color bins, this framework can easily be updated as additional image data is acquired, adapting to the variations in image acquisition such as lighting and the types of lesions that may be observed in particular geographic locations.
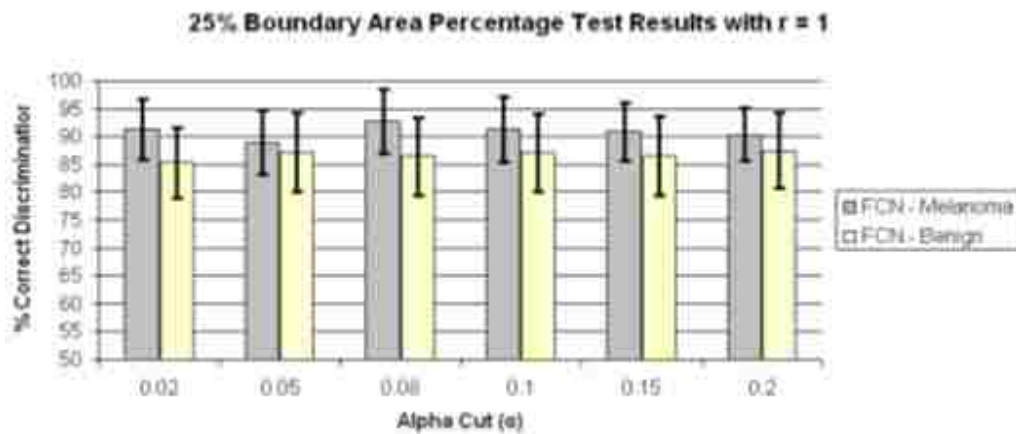
(a)

Figure 6. Average and Standard Deviation Test Results Over 25 Randomly Generated Training/Test Sets for 25% Boundary Area Percentage for $r = 1$ and $\alpha = 0.02, 0.05, 0.08,$ $0.1, 0.15$ and 0.2. FCN Denotes the Fuzzy Ratio Features Computed Using the Benign Training Images to Determine the Fuzzy Set B

Overall, $R(\alpha, r)$ yields higher true positive results for $r = 1$ than for $r = 0$ for all corresponding boundary area percentage cases and alpha-cuts, indicating that quantifying the grouping of similar colors provides more significant discrimination information than quantifying isolated color information. Experimental results also show that the outer 25% of the lesion enhances discrimination capability. Previous research with color feature

analysis in clinical images showed that the outer 10% of the skin lesion yielded the highest melanoma discrimination capability [26]. Finally, the fuzzy logic-based approach provided higher discrimination (true positive and true negative results) when computing $R(\alpha, r)$ based on determining **B** using the benign training set of images over the melanoma training set of images. This finding may be attributed to the presence of specified color clusters in benign lesions more consistently than in melanoma lesions. The features computed using the melanoma and benign training sets of images provide complementary information because only the specified class of training lesion images is used to populate the relative color histogram to determine **B**. Experimental results presented here provide further confirmation of the effectiveness of relative color, since the image sets were acquired from different patient populations via significantly different imaging systems.

The results found in the current study have relevance for clinicians. First, our results confirm results from a previous study that found that white or hypopigmented areas in the outer three deciles of the lesion were needed for the best discrimination of melanoma vs. benign [38]. Second, we have confirmed the clinical impression that colors within the lesion periphery, specifically the outer lesion deciles, are important in describing benign lesions, since a circumferential "hazy border", with coloration intermediate between that of skin and lesion, implies a benign lesion [32] and a "pink rim" is a sign of a melanoma [39]. Third, the patches or clusters of color that matter most are larger than a single pixel but still relatively small, in the single-digit range, as in the "peppering" of early malignancies [40, 41]. Thus a fuzzy logic-based description of lesion colors offers relevance to clinical descriptions of malignant melanoma.

5. CONCLUSIONS

In this research, a fuzzy logic-based color analysis approach was investigated for differentiating benign skin lesions from melanomas in dermoscopy images. The proposed approach extends previous research using a fuzzy clustering technique based on a trapezoidal

membership function to represent relative skin lesion color for a specified class of lesions and the degree to which relative skin lesion colors are from the specified class. Color histogram analysis on a training set of images for a specified class provided the basis for skin lesion color membership assignment. A fuzzy ratio for $r = 0$ and 1 was computed to reflect the area of the skin lesion with a specified degree of association to specified class color to the overall area of the lesion with any degree of color for that specified class. An automated thresholding technique, using the intersection between true negative and positive rates, was developed to differentiate benign skin lesions from melanomas based on the computed ratios over a training set of images. The thresholding determined from the training data is applied to the test data. Experimental results (as summarized in Figures 4 - 6) demonstrate that this technique can be used to successfully differentiate benign skin lesions from melanomas with true positive rates as high as 92.6% for the outer 25% of the lesion with $r = 1$ and $\alpha = 0.08$ based on the benign set of training images to determine **B**. These results provide a substantial improvement over the baseline fuzzy ratio feature for $r = 0$.

Future work on fuzzy logic-based color analysis will extend this research to larger data sets and explore combinations of percentile-based features, e.g., decile-based calculated features, as well as features based on various radii.

Author Contributions: R.J.S. and H.A.A. conceived and designed the experiments; H.A.A. and R.J.S. performed the experiments; R.J.S., H.A.A., W.V.S. and R.H.M. analyzed the data; R.J.S., H.A.A., W.V.S. and R.H.M. wrote the paper and R.J.S., H.A.A.,W.V.S. and R.H.M. revised the paper.

Conflicts of Interest: The authors declare no conflict of interest.

REFERENCES

[1] Marghoob, A. A. and Scope, A., 'The complexity of diagnosing melanoma.' The Journal of Investigative Dermatology, 2009, **129**(1), pp. 11–3, ISSN 1523-1747, doi: 10.1038/jid.2008.388.

[2] Stoecker, W. V. and Stolz, W., 'Dermoscopy and the Diagnostic Challenge of Amelanotic and Hypomelanotic Melanoma,' Archives of Dermatology, 2008, **144**(9), pp. 1207–10, ISSN 0003-987X, doi:10.1001/archderm.144.9.1207.

[3] Menzies, S. W., Moloney, F. J., Byth, K., Avramidis, M., Argenziano, G., Zalaudek, I., Braun, R. P., Malvehy, J., Puig, S., Rabinovitz, H. S., Oliviero, M., Cabo, H., Bono, R., Pizzichetta, M. A., Claeson, M., Gaffney, D. C., Soyer, H. P., Stanganelli, I., Scolyer, R. A., Guitera, P., Kelly, J., McCurdy, O., Llambrich, A., Marghoob, A. A., Zaballos, P., Kirchesch, H. M., Piccolo, D., Bowling, J., Thomas, L., Terstappen, K., Tanaka, M., Pellacani, G., Pagnanelli, G., Ghigliotti, G., Ortega, B. C., Crafter, G., Ortiz, A. M. P., Tromme, I., Karaarslan, I. K., Ozdemir, F., Tam, A., Landi, C., Norton, P., Kaçar, N., Rudnicka, L., Slowinska, M., Simionescu, O., Di Stefani, A., Coates, E., and Kreusch, J., 'Dermoscopic Evaluation of Nodular Melanoma,' JAMA Dermatology, 2013, **149**(6), p. 699, ISSN 2168-6068, doi:10.1001/jamadermatol.2013.2466.

[4] Soon, S. L., Solomon, A. R., Papadopoulos, D., Murray, D. R., McAlpine, B., and Washington, C. V., 'Acral lentiginous melanoma mimicking benign disease: The Emory experience,' Journal of the American Academy of Dermatology, 2003, **48**(2 SUPPL.), pp. 183–188, ISSN 01909622, doi:10.1067/mjd.2003.63.

[5] Kittler, H., Pehamberger, H., Wolff, K., Binder, M., and Levenstein, M., 'Diagnostic accuracy of dermoscopy.' The Lancet. Oncology, 2002, **3**(3), pp. 159–65, ISSN 1470-2045, doi:10.1016/S1470-2045(02)00679-4.

[6] Vestergaard, M. E., Macaskill, P., Holt, P. E., and Menzies, S. W., 'Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: A meta-analysis of studies performed in a clinical setting,' British Journal of Dermatology, 2008, **159**(3), pp. 669–676, ISSN 00070963, doi:10.1111/j.1365-2133.2008.08713.x.

[7] Ferris, L. K., Harkes, J. A., Gilbert, B., Winger, D. G., Golubets, K., Akilov, O., and Satyanarayanan, M., 'Computer-aided classification of melanocytic lesions using dermoscopic images,' Journal of the American Academy of Dermatology, 2015, **73**(5), pp. 769–776, ISSN 10976787, doi:10.1016/j.jaad.2015.07.028.

[8] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., and Thrun, S., 'Dermatologist-level classification of skin cancer with deep neural networks,' Nature, 2017, **542**(7639), pp. 115–118, ISSN 0028-0836, doi:10.1038/nature21056.

[9] Carrera, C., Marchetti, M. A., Dusza, S. W., Argenziano, G., Braun, R. P., Halpern, A. C., Jaimes, N., Kittler, H. J., Malvehy, J., Menzies, S. W., Pellacani, G., Puig, S., Rabinovitz, H. S., Scope, A., Soyer, H. P., Stolz, W., Hofmann-Wellenhof, R., Zalaudek, I., and Marghoob, A. A., 'Validity and Reliability of Dermoscopic Criteria Used to Differentiate Nevi From Melanoma,' JAMA Dermatology, 2016, **152**(7), p. 798, ISSN 2168-6068, doi:10.1001/jamadermatol.2016.0624.

[10] Argenziano, G., Puig, S., Zalaudek, I., Sera, F., Corona, R., Alsina, M., Barbato, F., Carrera, C., Ferrara, G., Guilabert, A., Massi, D., Moreno-Romero, J. A., Muñoz-Santos, C., Petrillo, G., Segura, S., Soyer, H. P., Zanchini, R., and Malvehy, J., 'Dermoscopy improves accuracy of primary care physicians to triage lesions suggestive of skin cancer,' Journal of Clinical Oncology, 2006, **24**(12), pp. 1877–82, ISSN 1527-7755, doi:10.1200/JCO.2005.05.0864.

[11] Siegel, R. L., Miller, K. D., and Jemal, A., 'Cancer statistics, 2016,' CA: A Cancer Journal for Clinicians, 2016, **66**(1), pp. 7–30, ISSN 00079235, doi:10.3322/caac.21332.

[12] Siegel, R. L., Miller, K. D., and Jemal, A., 'Cancer statistics, 2017,' CA: A Cancer Journal for Clinicians, 2017, **67**(1), pp. 7–30, ISSN 00079235, doi:10.3322/caac.21387.

[13] Bajaj, S., Marchetti, M. A., Navarrete-Dechent, C., Dusza, S. W., Kose, K., and Marghoob, A. A., 'The Role of Color and Morphologic Characteristics in Dermoscopic Diagnosis,' JAMA Dermatology, 2016, **152**(6), p. 676, ISSN 2168-6068, doi:10.1001/jamadermatol.2016.0270.

[14] Friedman, R. J., Rigel, D. S., and Kopf, A. W., 'Early detection of malignant melanoma: the role of physician examination and self-examination of the skin.' CA: a cancer journal for clinicians, 1985, **35**(3), pp. 130–151, ISSN 0007-9235, doi:10.3322/canjclin.35.3.130.

[15] Rubegni, P., Feci, L., Nami, N., Burroni, M., Taddeucci, P., Miracco, C., Munezero Butorano, M. A. G., Fimiani, M., and Cevenini, G., 'Computer-assisted melanoma diagnosis: a new integrated system.' Melanoma research, 2015, **25**(6), pp. 537–42, ISSN 1473-5636, doi:10.1097/CMR.0000000000000209.

[16] Andreassi, L., Perotti, R., Rubegni, P., Burroni, M., Cevenini, G., Biagioli, M., Taddeucci, P., Dell'Eva, G., and Barbini, P., 'Digital dermoscopy analysis for the differentiation of atypical nevi and early melanoma: a new quantitative semiology,' Archives of Dermatology, 1999, **135**(12), pp. 1459–65, ISSN 0003-987X.

[17] Landau, M., Matz, H., Ethel, T., Dvir, M., and Brenner, S., 'Computerized system to enhance the clinical diagnosis of pigmented cutaneous malignancies,' International Journal of Dermatology, 1999, **38**(6), pp. 443–446, ISSN 00119059, doi:10.1046/j.1365-4362.1999.00629.x.

[18] Umbaugh, S. E., Moss, R. H., and Stoecker, W. V., 'Automatic Color Segmentation of Images with Application to Detection of Variegated Coloring in Skin Tumors,' IEEE Engineering in Medicine and Biology Magazine, 1989, **8**(4), pp. 43–50, ISSN 07395175, doi:10.1109/51.45955.

[19] Green, A., Martin, N., Pfitzner, J., O'Rourke, M., and Knight, N., 'Computer image analysis in the diagnosis of melanoma.' Journal of the American Academy of Dermatology, 1994, **31**(6), pp. 958–64, ISSN 0190-9622.

[20] Seidenari, S., Burroni, M., Dell'Eva, G., Pepe, P., and Belletti, B., 'Computerized evaluation of pigmented skin lesion images recorded by a videomicroscope: comparison between polarizing mode observation and oil/slide mode observation,' Skin Research and Technology, 1995, **1**(4), pp. 187–191, ISSN 0909-752X, doi: 10.1111/j.1600-0846.1995.tb00042.x.

[21] Aitken, J. F., Pfitzner, J., Battistutta, D., O'Rourke, P. K., Green, A. C., and Martin, N. G., 'Reliability of computer image analysis of pigmented skin lesions of Australian adolescents,' Cancer, 1996, **78**(2), pp. 252–257, ISSN 0008-543X, doi:10.1002/(SICI) 1097-0142(19960715)78:2<252::AID-CNCR10>3.0.CO;2-V.

[22] Ercal, F., Chawla, A., Stoecker, W. V., Lee, H. C., and Moss, R. H., 'Neural network diagnosis of malignant melanoma from color images,' IEEE Transactions on Biomedical Engineering, 1994, **41**(9), pp. 837–845, ISSN 00189294, doi:10.1109/10.312091.

[23] Ganster, H., Pinz, P., Rohrer, R., Wildling, E., Binder, M., and Kittler, H., 'Automated melanoma recognition,' IEEE Transactions on Medical Imaging, 2001, **20**(3), pp. 233–239, ISSN 02780062, doi:10.1109/42.918473.

[24] Heckbert, P., 'Color image quantization for frame buffer display,' in 'Proceedings of the 9th annual conference on Computer graphics and interactive techniques - SIGGRAPH '82,' volume 16, ACM Press, New York, New York, USA, ISBN 0897910761, ISSN 0097-8930, 1982 pp. 297–307, doi:10.1145/800064.801294.

[25] Faziloglu, Y., Stanley, R. J., Moss, R. H., Van Stoecker, W., and McLean, R. P., 'Colour histogram analysis for melanoma discrimination in clinical images,' Skin Research and Technology, 2003, **9**(2), pp. 147–56, ISSN 0909-752X, doi:10.1034/j. 1600-0846.2003.00030.x.

[26] Chen, J., Stanley, R. J., Moss, R. H., and Van Stoecker, W., 'Colour analysis of skin lesion regions for melanoma discrimination in clinical images,' Skin Research and Technology, 2003, **9**(2), pp. 94–104, ISSN 0909-752X, doi:10.1034/j.1600-0846. 2003.00024.x.

[27] Stanley, R. J., Stoecker, W. V., and Moss, R. H., 'A relative color approach to color discrimination for malignant melanoma detection in dermoscopy images,' Skin Research and Technology, 2007, **13**(1), pp. 62–72.

[28] Stanley, R. J., Moss, R. H., Van Stoecker, W., and Aggawal, C., 'A fuzzy-based histogram analysis technique for skin lesion discrimination in dermatology clinical images,' Computerized Medical Imaging and Graphics, 2003, **27**(5), pp. 387–396, ISSN 08956111, doi:10.1016/S0895-6111(03)00030-2.

[29] Barata, C., Ruela, M., Francisco, M., Mendonca, T., and Marques, J. S., 'Two systems for the detection of melanomas in dermoscopy images using texture and color features,' IEEE Systems Journal, 2014, **8**(3), pp. 965–979, ISSN 19379234, doi:10.1109/JSYST.2013.2271540.

[30] Stoecker, W. V., *Computer applications in dermatology*, Igaku-Shoin New York, 1993.

[31] Nie, D., 'Classification of melanoma and clark nevus skin lesions based on Medical Image Processing Techniques,' in 'ICCRD2011 - 2011 3rd International Conference on Computer Research and Development,' volume 3, IEEE, ISBN 9781612848372, 2011 pp. 31–34, doi:10.1109/ICCRD.2011.5764239.

[32] Argenziano, G., Soyer, H. P., De Giorgi, V., Piccolo, D., Carli, P., and Delfino, M., 'Interactive atlas of dermoscopy (Book and CD-ROM),' 2000.

[33] McLean, R. P., *Tumor classification based on relative color analysis of melanoma and non-melanoma tumor images*, Ph.D. thesis, 1994.

[34] Stoecker, W. V., Li, W. W., and Moss, R. H., 'Automatic detection of asymmetry in skin tumors,' Computerized Medical Imaging and Graphics, 1992, **16**(3), pp. 191–197, ISSN 08956111, doi:10.1016/0895-6111(92)90073-I.

[35] Hance, G., Umbaugh, S., Moss, R., and Stoecker, W., 'Unsupervised color image segmentation: with application to skin tumor borders,' IEEE Engineering in Medicine and Biology Magazine, 1996, **15**(1), pp. 104–111, ISSN 07395175, doi:10.1109/51.482850.

[36] KLIR, G. J. A. and FOLGER, T. A. A., *Fuzzy Sets, Uncertainty, and Information*, Prentice Hall, 1988, ISBN 9780133459845.

[37] Joe Stanley, R., Stoecker, W. V., Moss, R. H., Rabinovitz, H. S., Cognetta, A. B., Argenziano, G., and Peter Soyer, H., 'A basis function feature-based approach for skin lesion discrimination in dermatology dermoscopy images,' Skin Research and Technology, 2008, **14**(4), pp. 425–435, ISSN 0909752X, doi:10.1111/j.1600-0846.2008.00307.x.

[38] Dalal, A., Moss, R. H., Stanley, R. J., Stoecker, W. V., Gupta, K., Calcara, D. A., Xu, J., Shrestha, B., Drugge, R., Malters, J. M., and Perry, L. A., 'Concentric decile segmentation of white and hypopigmented areas in dermoscopy images of skin lesions allows discrimination of malignant melanoma,' Computerized Medical Imaging and Graphics, 2011, **35**(2), pp. 148–154, ISSN 08956111, doi:10.1016/j.compmedimag.2010.09.009.

[39] Rader, R. K., Payne, K. S., Guntupalli, U., Rabinovitz, H. S., Oliviero, M. C., Drugge, R. J., Malters, J. J., and Stoecker, W. V., 'The Pink Rim Sign : Location of Pink as an Indicator of Melanoma in Dermoscopic Images,' Journal of skin cancer, 2014, **2014**, p. 719740, ISSN 2090-2905, doi:10.1155/2014/719740.

[40] Braun, R. P., Gaide, O., Oliviero, M., Kopf, A. W., French, L. E., Saurat, J. H., and Rabinovitz, H. S., 'The significance of multiple blue-grey dots (granularity) for the dermoscopic diagnosis of melanoma,' British Journal of Dermatology, 2007, **157**(5), pp. 907–913, ISSN 00070963, doi:10.1111/j.1365-2133.2007.08145.x.

[41] Stoecker, W. V., Wronkiewiecz, M., Chowdhury, R., Stanley, R. J., Xu, J., Bangert, A., Shrestha, B., Calcara, D. A., Rabinovitz, H. S., Oliviero, M., Ahmed, F., Perry, L. A., and Drugge, R., 'Detection of granularity in dermoscopy images of malignant melanoma using color and texture features,' Computerized Medical Imaging and Graphics, 2011, **35**(2), pp. 144–147, ISSN 08956111, doi:10.1016/j.compmedimag.2010.09.005.

II. ALGORITHM ENHANCEMENTS FOR IMPROVEMENT OF LOCALIZED CLASSIFICATION OF UTERINE CERVICAL CANCER DIGITAL HISTOLOGY IMAGES

Haidar A. Almubarak[1], Peng Guo[1], R. Joe Stanley[1], Rodney Long[2] , Sameer Antani[2], George Thoma[2], Rosemary Zuna[3], Shelliane R. Frazier[4], Randy H. Moss[1] William V. Stoecker[5], and Jason Hagerty[5]

[1]Department of Electrical & Computer Engineering

Missouri University of Science and Technology Rolla, Missouri 65409–0050

Tel: 573–341–6896

[2]Lister Hill National Center for Biomedical Communications for National Library of Medicine, National Institutes of Health, DHHS in Bethesda, MD, USA

[5]Stoecker and Associates, Rolla, MO

Email: h.almubarak@ieee.org, {pgp49,stanleyj, rhm, wvs}@mst.edu, hagerty.jason@gmail.com, {rlong,antani,gthoma}@mail.nih.gov

ABSTRACT

In prior research, we introduced an automated, localized, fusion-based approach for classifying squamous epithelium into Normal, CIN1, CIN2, and CIN3 grades of cervical intraepithelial neoplasia (CIN) from digitized histology image analysis. The image analysis approach partitioned the epithelium along the medial axis into ten vertical segments. Texture, cellularity, nuclear characterization and distribution, and acellular features were computed from each vertical segment. The individual vertical segments were CIN classified, and the individual classifications were fused to generate an image-based CIN assessment. In this research, image analysis techniques are investigated to improve the execution time of the algorithms and the CIN classification accuracy of the baseline algorithms. For an

experimental data set of 117 digitized histology images, execution time for exact grade CIN classification accuracy was improved by 32.32 seconds without loss of exact grade CIN classification accuracy (80.34% vs. 79.49% previously reported) for this same data set

1. INTRODUCTION

There were 527,642 new invasive cervical cancer cases reported in 2012, with an estimated 265,672 deaths [1]. Around 84% of the new cases were reported in less developed countries. Screening for cervical cancer and its precursor lesions is carried out using a Papanicolaou (Pap) test. Definitive evaluation requires histology slides of biopsied cervical tissue; interpretation of these slides is done by an expert pathologist [2]. Pathologists seek to detect cervical intraepithelial neoplasia (CIN), which is a pre-malignant condition for cervical cancer. A cervical biopsy is classified as normal (no CIN) or one of three CIN grades: CIN1 (mild dysplasia), CIN2 (moderate dysplasia), or CIN3 (severe dysplasia) by identifying the atypical cells in the epithelium by the visual inspection of histology slides [3, 4, 5]. Figure 1 shows examples of the CIN grades. Delayed maturation with an increase in immature atypical cells from bottom to top of the epithelium has been observed as CIN increases in severity [6]. As shown in Figure 1, atypical immature cells are often present in the bottom third of the epithelium for CIN1 (Figure 1b). The atypical immature cells commonly appear in the bottom two thirds of the epithelium for CIN2 (Figure 1c), and atypical immature cells are typically present in the full thickness of the epithelium for CIN3 (Figure 1d). In addition to analyzing the progressively increasing quantity of atypical cells from bottom to top of the epithelium, identification of nuclei atypia is also significant [6]. Nuclei atypia is characterized by variation in nuclei shapes and sizes. Visual assessment of nuclei differences is challenging, contributing significantly to diagnostic inter- and intra-pathologist variation. Nuclei deformation characterization

and atypical cell quantization play a critical role in identifying CIN grade [3]. Computer-assisted CIN diagnostic methods have been explored in other studies [5, 7, 8, 9, 10, 11]. These methods studied texture features [10], nuclei detection and Delaunay triangulation analysis [8, 11], medial axis determination , and localized CIN grade assessment [5, 7, 8].
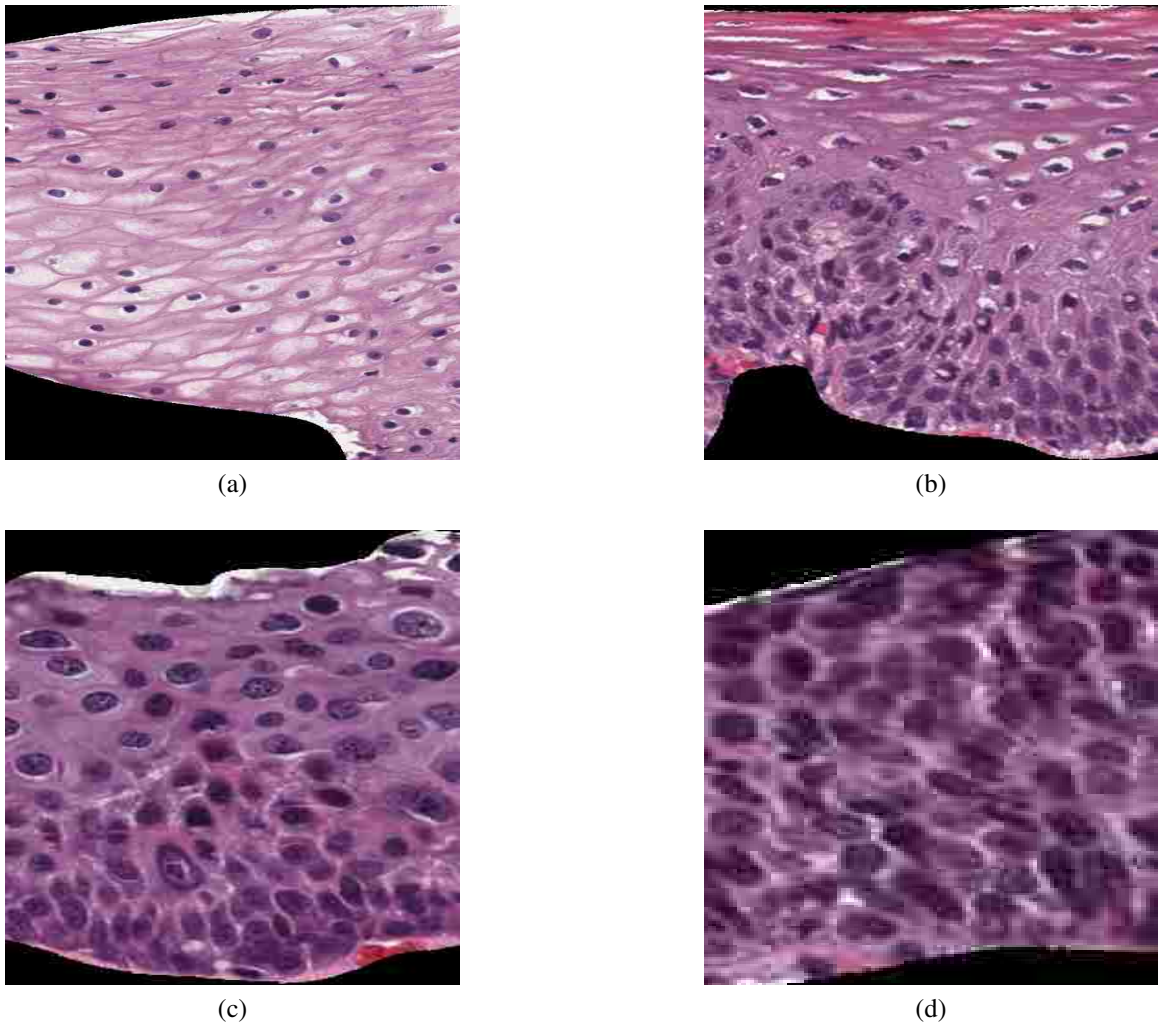
Figure 1. CIN Grades Example (a) Normal, (b) CN1, (c) CIN2, (d) CIN3

In previous research, we investigated a localized fusion-based approach to classify epithelium into four different CIN grades [7, 8] . The approach described in [7] yielded 88.5% accuracy for an image set of 61 digitized histology images. For this experimental dataset, the computation time was high, with an average feature extraction time of 36.18 seconds per image; some images took more than 200 seconds. Thus, it would be impractical

to utilize those image analysis techniques in a semi-automated tool in a clinical setting to assist expert pathologists in their clinical assessment of digitized histology images The goal of this research is to improve the computation time for the image analysis approach used in [7] without degrading the classification accuracy. In this research, we have examined 117 cervical epithelium images that were manually cropped out of the digitized histology slides of Hematoxylin and Eosinophil (H&E) preparations of tissue sections. An expert pathologist then created "ground truth" diagnostic data by classifying the images into one of four grades (Normal, CIN1, CIN2, and CIN3). In this paper we investigate image analysis algorithm modifications to decrease computation time for automatic CIN classification of these images, as described in the methods section, and, for each modification, we calculate the execution time and determine the CIN discrimination accuracy. Changing the region of interest (ROI) has been used before [12] to optimize performance of CAD systems. It was shown that increasing the ROI will increase processing time but not necessarily the accuracy [12]. Here we are not changing the region of interest, but changing other aspects of the feature extraction algorithms to reduce computation time. Section 2 of this article presents the methods used. Section 3 describes the experiments performed, and Section 4 presents the results analyses.

2. METHOD

In this paper we are utilizing 117 digitized histology images in extending the study from [7, 8]. The process includes the steps: 1) Locate the medial axis of the segmented epithelium; 2) Divide the epithelium into vertical segments; 3) Extract features from the individual vertical segments; 3) classify each vertical segment; and 4) Fuse CIN grades of the vertical segments to obtain a CIN grade for the whole epithelium image. Detailed descriptions of the algorithm implementations are given in [7].

2.1. Overview of Medial Axis Detection and Vertical Segment Partitioning Algorithms. A summary of the approach to detect the medial axis and to partition the epithelium to obtain vertical segments follows. First, the medial axis for the interior 60% of the whole epithelium sample is determined using the distance transform as described in [8]. The leftmost and rightmost 20% of the sample (L-R 20%) use the method described in [7]. Second, the image and its corresponding epithelium binary mask are read; then the mask is applied to the image. The rotation angle is determined and used to rotate the image and the mask. Third, the bounding box of the rotated binary mask corresponding to the epithelium is used in the distance transform function. For the interior 60% the highest intensity pixels of the distance transform along vertical slices (max of each column) are used as the medial axis. For the L-R 20%, the center line of the bounding-box that divides the epithelium into top and bottom halves is calculated. Then the intersection points between the center line and the edge of the 60% interior medial axis are determined; these will be the starting points for medial axis correction. Fourth, the edges of the epithelium mask are computed using the Sobel filter; these edges are divided into 20%, 60%, and 20% parts. The median points of the left and right edges of the entire segment are used as endpoints of the medial axis. The medial axis in the L-R 20% of the segments is then rotated between the intersection points and the end of the epithelium to align the ends with the median points found in the edges. The details of this process can be found in [7] and [8]. Figure 2 presents an image example of the epithelium automatically determined using the algorithm from [8] overviewed above. Figure 2 also shows the partitioning of the epithelium into 10 vertical segments, where each segment's midpoint is perpendicular to the midpoint of the segment's medial axis.

2.2. Feature Extraction from Individual Vertical Segment Images. From [7], twenty-seven features were extracted from each individual vertical segment for CIN classification. The features are divided into six groups: a) Texture Features (F1-F10), b) Cellular
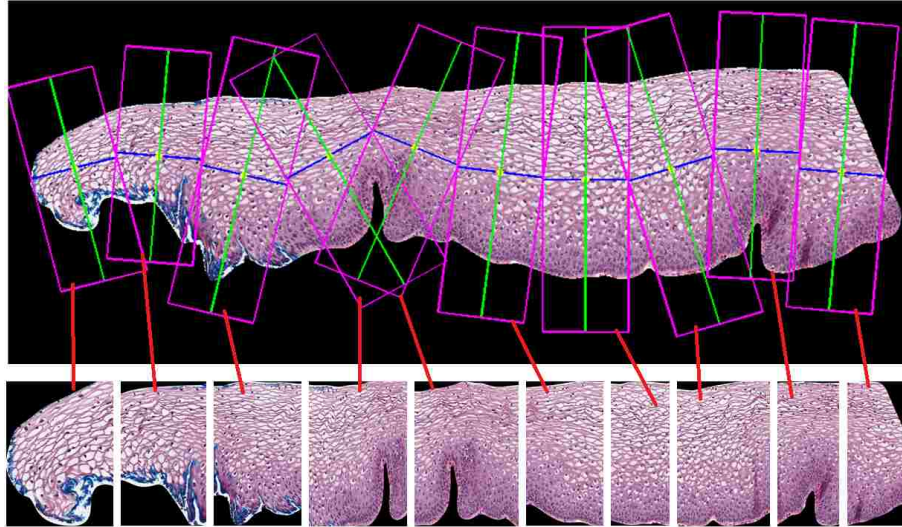
Figure 2. Vertical Segments of the Epithelium

feature (F11-F13), c) Nuclei Features (F14,F15), d) Acellular features (F16-F22), e) Combined features (F23,F24), f) Triangular features (F25-F27). The feature extraction process is described below.

2.2.1. Texture features. A grayscale luminance copy of the segment was used to derive first-order structural measures (contrast, energy, correlation, and uniformity). The gray-level co-occurrence matrix (GLCM) [5] [7] was computed from the grayscale image and four second-order measures (contrast, energy, correlation, and uniformity) were calculated from the GLCM. These features are denoted as F1-F10: Contrast of segment (F1), Energy of segment (F2), Correlation of segments (F3), Segment homogeneity (F4), Contrast of GLMC in horizontal (F5) and vertical (F6) directions, Correlation of GLMC in horizontal (F7) and vertical (F8) directions, Energy of GLMC in horizontal (F9) and vertical (F10) directions.

2.2.2. Cellular features. The luminance image Figure 3 shows three ranges of intensities corresponding to acellular, nuclei, and cytoplasm areas
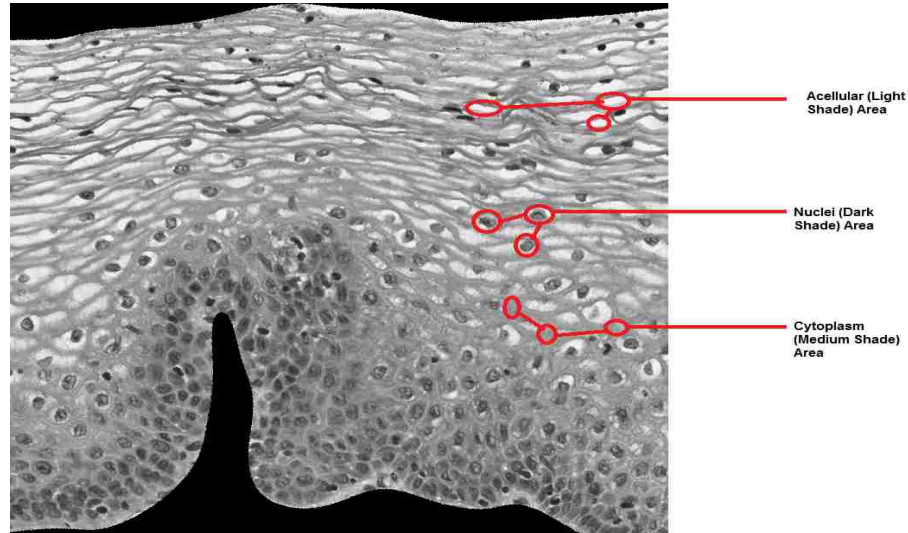
Figure 3. Cellular Features Example

The K-Means clustering algorithm with K =3 was used to cluster all pixels in the luminance image [8] into 3 clusters: light cluster, medium cluster, and dark cluster. The ratio of the pixels in each cluster to the total pixels in the epithelium (Features F11-F13) is calculated as follow:

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

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

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

2.2.3. Nuclei features. The nuclei ratio or dark areas discussed above only compute the percentage of epithelium considered to be nuclei but does not quantify the number of nuclei. From Figure 1 it can be seen that the number and shape of nuclei is correlated with the degree of CIN–the more nuclei we have, the higher the CIN grade. An algorithm was developed to detect and quantify the nuclei [7] to facilitate better classification precision.

The process is performed twice to generate two masks that are added to generate the final mask. The second time the same process is carried out except after sharpening the image in the preprocessing stage the background is inverted. Inverting the background helps in detecting nuclei that are not detected in the first run. Figure 4 shows example of nuclei being detected with inverted and non-inverted background images.

Nuclei detection consists of a preprocessing stage and a detection stage. The preprocessing steps are as follows: The segment needs to be converted first into a grayscale image I, then an averaging filter (A) is applied to I, given in Equation 4 as:

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

After the grayscale image is filtered, it is then sharpened to emphasis the dark shading areas (Equation 5) where k is empirically determined as 2.25:

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

Second, the Matlab function imclose is then used with a circular structuring element of radius 4 to perform morphological closing on the nuclei mask image. The resultant image will still contain some holes that will be filled using Matlab's *imfill* function. After filling the holes, Matlab's *imopen* function is used to perform morphological opening with a circular structuring element of radius 4. Small areas corresponding to noise objects (non-nuclei objects) within the epithelium region of interest from the mask are then eliminated with the area opening operation using the Matlab function *bwareaopen*. The execution time bottleneck lies in the background inversion and clustering algorithms. To preserve the edges of the epithelium, the inversion process uses a sliding window of size 2 pixels; if the window is not on the edge, it will invert the pixels covered by the window, otherwise keep it intact. For a segment of size $M * N$ it will take $\frac{M*N}{2}$ operations. The clustering works by sliding

a non-overlapping window of size 29 by 29 pixels and clustering the values covered by this window. Localizing the process this way enhances the detection of nuclei especially for smaller images. The number of times the clustering algorithm runs is given by $\left\lceil \frac{M}{wh} \right\rceil * \left\lceil \frac{N}{wl} \right\rceil$ where M and N are image dimensions and wh and wl are window dimensions. With small window size and large image size, the processing time for larger images is relatively high compared to a larger window size. Using the final nuclei mask nuclei features F14 and F15
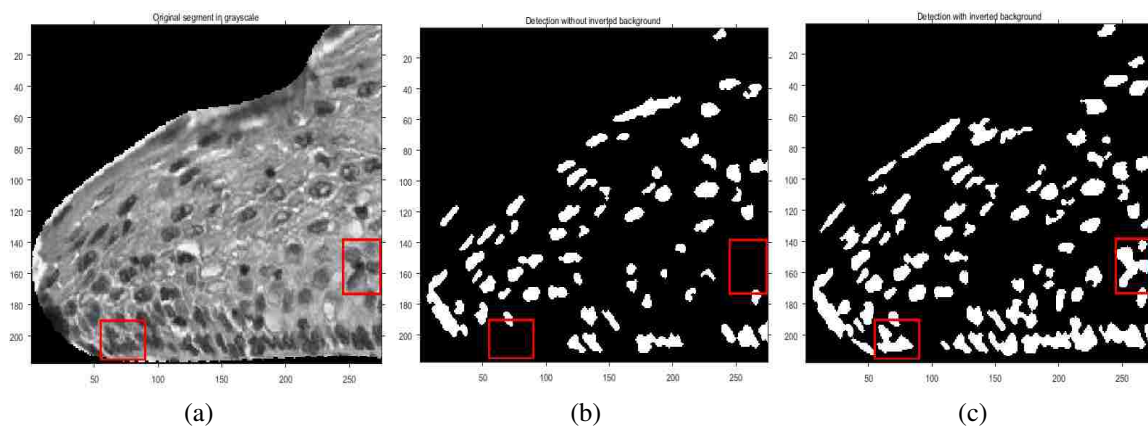
Figure 4. Effect of Inverting Background on the Detection of Nuclei. (a) Shows the Segments Grayscale Version, (b) Shows the Nuclei Being Detected Without Inverting and (c) Shows the Nuclei Detected After Inverting

are calculated as follows:

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

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

Where "nuclei area total" represents the number of pixels in the white area, "nuclei number" is the total number of white regions, and "total non-nuclei area" is the sum of black pixels within the epithelium region.

2.2.4. Acellular features. Acellular area detection is similar to nuclei detection. This consists of two steps: first detect the light area and create a binary mask of it; then use the mask in the R, G, B, and Luminance planes to calculate the features. The light area detection has the following steps: Step 1: Convert the original image from RGB color space to L*a*b* color space, then select the luminance component L* Step 2: Perform adaptive histogram equalization on the image from step 1 using Matlab's *adapthisteq*. Step 3: Apply a threshold of 0.6 to eliminate the dark nuclei regions and to retain the lighter nuclei and epithelium along with the light areas. Step 4: Segment the light areas using the K-means algorithm with K equal to 4. The K-means algorithm input is the histogram-equalized image from step 2 multiplied by the binary image from step 3. Step 5: Remove from the image all objects having an area less than 100 pixels, determined empirically, using the Matlab function *regionprops*. A morphological closing is performed with a disk structure element of radius 2. After generating the final mask in step 5 above the acellular features F16-F22 are calculated as follows:

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

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

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

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

$$Luminance\ ratio = \frac{Light\ Area\ luminance}{Background\ luminance} \tag{12}$$

$$Ratio\ light\ to\ segment\ area = \frac{Light\,Area}{Segment\ Area} \qquad (13)$$

$$Ratio\ light\ to\ background\ area = \frac{Light\,Area}{Background\ Area} \qquad (14)$$

where Segment Area gives the epithelium area within the vertical segment; Light Area is the area of all light area regions; Light Number corresponds to the number of light areas; Background Area represents the total number of non-nuclei and non-light area pixels inside the epithelium within the vertical segment (i.e., background area); Light Area Intensity, Light Area Red, Light Area Green, Light Area Blue, and Light Area Luminance are the average intensity of red, green, blue, and luminance, respectively, inside epithelium of the vertical segment; Background Intensity, Background Red, Background Green, Background Blue, and Background Luminance are the average intensity of red, green, blue, and luminance, of the non-nuclei and non-light area pixels within the epithelium of the vertical segment.

2.2.5. Combined features. After the nuclei and the acellular features were extracted, two new features were calculated to capture the relative increase in nuclei numbers as CIN grade increases. These features are the ratio of the acellular number to the nuclei number (F23), and the ratio of the acellular area to the total nuclei area (F24). These features are calculated as follows:

$$Ratio\ acellular\ number\ to\ nuclei\ number = \frac{Light\ Number}{Nuclei\ Number} \qquad (15)$$

$$Ratio\ acellular\ area\ to\ nuclei\ area = \frac{Light\ Area}{Nuclei\ Area} \qquad (16)$$

Where Light Number and Nuclei Number represent the total number of light areas and nuclei respectively.

2.2.6. Triangular features. The centroids of the nuclei detected in Section 2.2.3 were used as vertices to generate Delaunay triangles. The vertical segment being processed is sub-divided into three horizontal layers. The aim is to associate the presence of increasing nuclei throughout the epithelium with increasing CIN grades. After locating the vertices for Delaunay triangles (DT), the DT algorithm iteratively selects point triples to become vertices of each new triangle created [5]. After the triangles in three different layers were obtained, the following features are computed: number of triangles in top layer, middle layer, and bottom layer, F25-F27, respectively [7].

2.3. Individual Vertical Segment and Image-based CIN Classification. After extracting the 27 features from all individual segments of the image, Support Vector Machine (SVM) and Linear Discriminant Analysis (LDA) classifiers are trained and tested using a leave-one-image out approach as described in [7]. The process goes as follow: One image (10 segments) is taken as the test image and the remaining images are used for the training. The features from the training set are then normalized using the mean and standard deviation. The normalized features are then used to train an SVM or LDA classifier using the expert (RZ) CIN labels for the whole image applied to each of the vertical segments. The trained model is then used to classify the individual segments of the test image, and the resulting predicted CIN grades are fused to obtain the CIN grade of the entire test epithelium image using a voting scheme by assigning the test image the most frequent class grade of the individual segments; if there is a tie assign the highest class grade (e.g. if there is a tie between CIN2 and CIN3 assign the image CIN3). The process is then repeated for all images as the test image.

2.4. Proposed Changes to Feature Extraction Algorithms to Improve Timing for Clinical Application. Using the epithelium image analysis and classification algorithms from [7], the baseline time to generate CIN grade output was determined for the experimental data set. The experimental data set of 117 images took an average of 49.88 seconds per image for preprocessing and extraction of 27 features with a maximum reaching 223.20

seconds based on the image analysis and classification algorithms used in [7]. A graphical user interface (GUI) application was developed in Matlab®, along with the image analysis and classification algorithms. The profiling tool in Matlab® [13] was used to determine which parts of the algorithm consume the greatest time in the feature extraction process. The features extraction workflow starts by reading an image along with its mask, the mask is then applied to the image; the image is then aligned correctly, by rotating major axis, for vertical segmentation. The medial axis is then detected and used to find the bounding box of each segment. Features are extracted from each segment by first cropping the segment using its bounding box. The texture, cellular, triangles, nuclei, and acellular features can be extracted independently, but combined features cannot be computed before extracting nuclei and acellular features As discussed in Section 2.2.3, there are three factors that lead to primary bottlenecks affecting the processing time for each image, including the image size, the window size used in detecting the nuclei, and inverting the background during the nuclei detecting preprocessing phase. The larger the image or the smaller the window size the more time it takes to perform the nuclei preprocessing operations. In order to mitigate this, we propose the following modifications to the process:

1. Resize the images to reduce resolution.

2. Change the window size inside the masking function.

3. Change the inversion process in the preprocessing step.

Referring to the workflow discussed above, resizing the image occurs after applying the mask and orienting the image but before finding the medial axis; this will affect all subsequent steps; hence, all features will be affected. Changing the window size for nuclei detection or background inversion will only affect the nuclei features (F14-F15) and combined features (F23-F24). Although most of the features are ratios, resizing the image will still affect them slightly.

We examined the proposed solutions with variations in window sizes and resizing process. For the resizing, two approaches have been examined. The first approach is to scale all images with a single scaling factor of 20%; the second approach is to scale the images dynamically based on the original size of the image. For the dynamic resizing, a baseline image that has a total of 785,841 pixels was used. The following criteria of resizing were used:

- If the image is larger than the baseline by less than 20%, do not resize it.

- If the difference between the baseline and image size is 20% to 45%, reduce the image by 20%

- If the difference between the baseline and image size is greater than 45%, reduce the image by 50%

The window sizes used, given as the length of window sides, are the original 29, 71, and 87 pixels. The baseline, resizing criteria, and window sizes were chosen empirically to preserve information while improving processing time. Using a large image as the baseline will keep most of the images at their original size, and this will not provide improvement in processing time. The average diameter of a normal human cell nucleus is about 6 μm [14]. With the 10x magnification used, a normal nucleus will be bounded by an area of 12 pixels by 12 pixels. As CIN grade progresses, the average size of the nuclei increases, the shape deforms, and the nuclei get closer to each other. We need a window size that can adapt to this deformation but at the same time isolate the nuclei if possible. Having a window size for nuclei detection that is too large will cause many nuclei to fall within the window and it will result in poor isolation of each nucleus. Figure 5 shows how the window size affects the resultant nuclei that were detected. For the background inversion, instead of using a sliding window, we inverted the whole background at once by changing all 0 pixels into 1s (fast background inversion). Using a larger window size will improve the speed but will result in errors in detected nuclei which will affect the classification result.
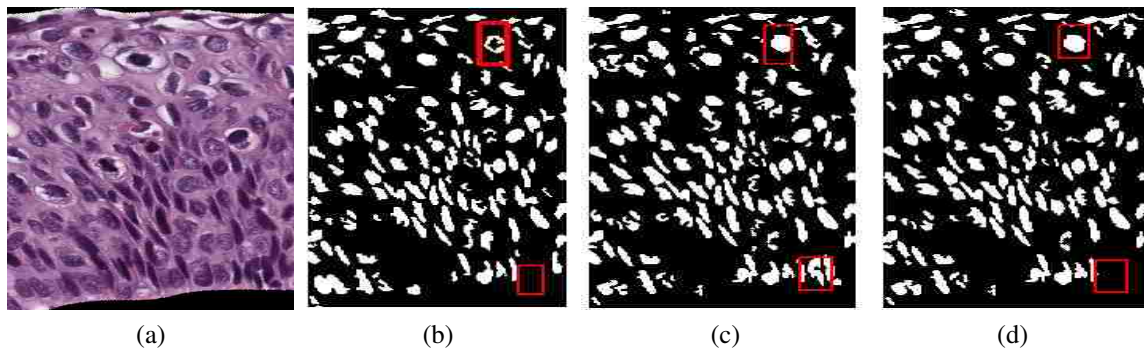
Figure 5. Effect of Window Size in Nuclei Detection. (a) Shows the Original Segment, (b) Shows the Nuclei Being Detected When 29*29 Window Size is Used, (c) Shows the Nuclei Being Detected with 87*87 Window Size, and (d) Shows the Nuclei Being Detected When Using 150*150 Window Size

2.5. Graphical User Interface Application. A Graphical User Interface (GUI) application was developed using Matlab (Figure 6) for clinical utilization as a diagnostic aide for CIN assessment of digitized histology images. The user starts by browsing for an image, then clicking analyze. The analyze button will invoke the feature extraction function; the feature extraction is based on the best performing algorithm, see Section 3. After the features are extracted, the small segments will be displayed and the features will be stored in memory for classification; the user can then click the classify button which will use the trained model, based on the classification type radio button selected, to classify the segments and the whole image. The whole image class and the segments classes' are displayed under each individual vertical segment. The application was provided to RZ for evaluation as diagnostic aide.

3. EXPERIMENTS AND RESULTS ANALYSIS

The dataset used in this research consists of 117 digitized histology images which were classified by an expert pathologist (RZ) as: 41 Normal, 23 CIN1, 22 CIN2, and 31 CIN3. Using all combinations of the proposed solutions in Section 2.4, there are
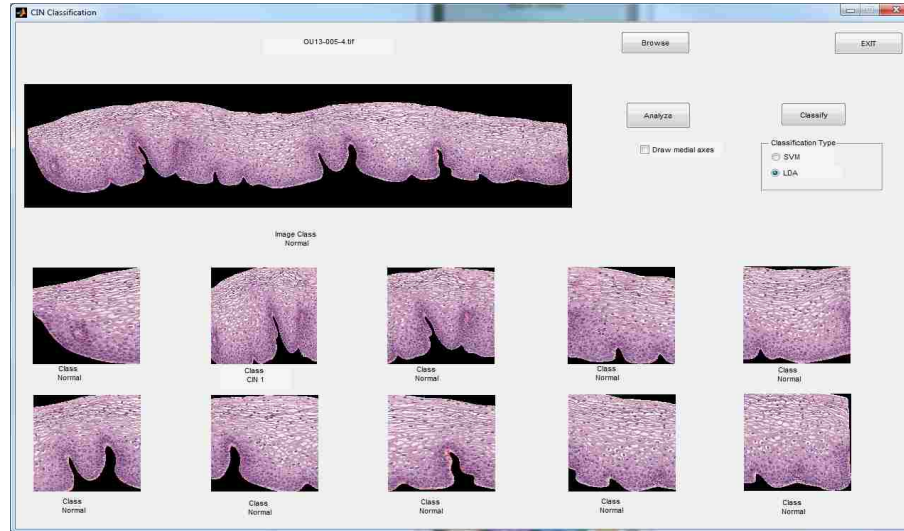
Figure 6. GUI for CIN Classification

20 total experiments. Algorithm execution times for the experiments performed in this study were based on a desktop computer with the following specifications: Windows 7 Enterprise SP1 64-bit operating system, Intel CORE i5-2400 3.1 GHz processor, 8 GB RAM, and Matlab® R2012a code development environment. Experiments 1-20 follow the same process described in [7] using a leave-one-image-out approach where one image (10 segments) was used for testing and the remaining 116 images used for training. Based on the process above the following experiments are performed:

1. Use the original code used in [7] for feature extraction. This acts as the baseline for the feature extraction time and classification accuracy.

2. Vectorize specific operations in the image analysis algorithms. The remaining experiments are based on the vectorized code.

3. Reduce image size by 20% and maintain the window size for nuclei detection at 29x29.

4. Dynamically resize images.

5. Use the original image size with the window size for nuclei detection at 71x71.

6. Use the 20% reduced size images with the window size for nuclei detection at 71x71.

7. Use the dynamically resized images with window size 71.

8. Use the original size image with the window size for nuclei detection at 87x87.

9. Use the window size for nuclei detection at 87x87

10. With image sizes reduced by 20%.

11. Use window size for nuclei detection at 87x87 with image sizes reduced dynamically.

12.-19. Repeat Experiments 2-10 with the modified algorithm for background inversion.

20. Compute 27 features extracted from the whole epithelium image without partitioning the images into vertical segments.

Table 1 summarizes the results of these experiments. The first column lists the different experiments performed The second column shows the background inversion used in each experiment: either the original algorithm, taking into account epithelium edges, or the modified algorithm, where edges were not taken into consideration when inverting the background. The third column indicates wither the image was resized in the experiment or the original size was used. Resizing can be either none when the original image sizes were used, fixed denoting a 20% resize or dynamic which denotes different resize based on criteria discussed in Section 2.4. The fourth column, Window size, represents the window used in nuclei detection. The fifth and sixth columns show accuracy of LDA and SVM algorithms classification for the exact class label; accuracy is compared with the expert pathologist classification. The last four columns show total time statistics for the time taken to extract the features from all 117 images, including the mean, standard deviation, maximum, and minimum time for each experiment.

From Table 1, there are several observations. First, the baseline experiment (Experiment 1) took an average time of 49.88 seconds to preprocess and extract the 27 features and resulted in 79.49% classification accuracy. Second, in Experiment 2, vectorization code

was used, which reduced the average time to process each image in the 117 experimental data set to 30.86 seconds, with a maximum time of 134.21 seconds. This modification reduces the time to process the images without changing extracted features; while this is a substantial reduction in time (from 49.88 seconds to 30.86 seconds), this time might still be far from being practical. We needed to examine the algorithm to find ways to improve the execution time. The profiling tool showed that detecting nuclei consumes most of the computation time, with an average of 9.96 seconds, which is approximately 30% of the processing time. Third, in experiments 5 and 8, we obtained the highest accuracy (81.20%); Experiment 8 (window size 87x87) had a slight time improvement (28.93 seconds) over Experiment 5 (window size 71x71), which took 29.10 seconds. Fourth, we can see from the table that just changing the window size does not improve the time by a big factor, but resizing the image improves the time significantly. The average time decreased by only 1.76 seconds when the window size changed from 29 (Experiment 2) to 71 (Experiment 5) and only 1.93 seconds was achieved when the window size changed to 87. Keeping all parameters fixed and changing the image size dynamically (Experiment 19) gives the best timing across all experiments. The best average time achieved was 12.15 seconds when using fast background inversion and dynamic resizing with window size 71 (Experiment 16). Using a larger window size will improve the time only slightly, with the biggest improvement coming from resizing the image. However, as we saw in Figure 5, the original window size missed some nuclei; the bigger window improves the accuracy, but using an even larger window size such as the 150x150 in Figure 5 will lead to other nuclei being undetected.

Fifth, the baseline accuracy (Experiment 1), using the original image sizes and original background inversion method with window size 29, the accuracy was only 79.49%. The best accuracy was 81.20% when the original image size was used with the original background inversion method and either 71x71 or 87x87 window size which took an average time of 29.10 and 28.93 seconds respectively to extract the features from Experiments 5 and 8, respectively. Using images resized by 20% and the new fast background inversion

with window size 71 gives the best balance between feature extraction time (17.56 seconds) and classification accuracy (80.34%) using the LDA classifier (Experiment 15). The computation time improved by 64% from 49.88 seconds to 17.56 seconds, and the accuracy is also improved slightly from 79.49% to 80.34%. This experiment also outperforms the results when using the whole image described in [7]. In Experiment 20 the accuracy was only 68.38% and the average feature extraction time was 71.35 seconds. Higher accuracy with less execution time enhances the user experience. We can also see that the LDA outperforms SVM in terms of classification except when using dynamically resized images. Sixth, we examined resizing the images, using different window sizes during nuclei detection, and using fast background inversion in the nuclei detection preprocessing step. Based on the results reported in Table 1 we achieved 74.49% improvement in the average feature extraction time (from 49.88 seconds to 12.72 seconds) at the expense of 5.13% decrease in accuracy from the baseline (79.49% to 74.37%); just by using vectorization, the average feature extraction time was reduced from 49.88 seconds to 30.86 seconds with the same accuracy. We obtained accuracy better than the baseline, 80.34%, with a simultaneous improvement in terms of execution time with average of 17.56 seconds.

The best exact label CIN classification results were obtained from Experiment 8 using the original background inversion algorithm with window size 87 for nuclei detection; this gives 81.20% classification accuracy with 28.93 seconds average execution time, which is better than the baseline 79.49% accuracy and 49.88 seconds execution time. Computer Aided Diagnosis (CAD) systems strive to give the best diagnosis accuracy but sometimes at the expense of more execution time which hinders the users' experience. In this study, we improved the accuracy of the baseline algorithm and reduced the execution time simultaneously.

Table 1. Summary of Experimental Results

Exp#	Background Inversion	Resizing	Window Size	Classification Accuracy		Run Time			
				LDA	SVM	AVG	STD	Max	Min
1	Original Alg.	None	29	79.49	73.50	49.88	44.09	223.19	9.13
2	Original Alg.	None	29	79.49	73.50	30.86	26.83	134.21	6.32
3	Original Alg.	Fixed	29	76.92	77.78	22.11	18.51	95.20	5.42
4	Original Alg.	Dynamic	29	68.38	72.65	14.90	12.50	90.37	6.16
5	Original Alg.	None	71	81.20	72.65	29.10	25.22	125.31	5.97
6	Original Alg.	Fixed	71	79.49	75.21	20.66	17.18	87.12	5.21
7	Original Alg.	Dynamic	71	70.09	73.50	14.27	11.51	79.07	6.02
8	Original Alg.	None	87	81.20	70.94	28.93	24.96	125.13	6.10
9	Original Alg.	Fixed	87	80.34	78.63	20.63	17.40	87.46	3.53
10	Original Alg.	Dynamic	87	68.38	73.50	14.18	11.53	80.44	5.96
11	Modified Alg.	None	29	76.92	73.50	26.91	23.02	115.00	5.93
12	Modified Alg.	Fixed	29	76.92	76.92	18.93	15.58	80.17	5.22
13	Modified Alg.	Dynamic	29	70.09	73.50	13.24	10.12	71.90	5.96
14	Modified Alg.	None	71	80.34	72.65	24.64	20.98	108.60	5.61
15	Modified Alg.	Fixed	71	80.34	76.07	17.56	14.09	73.81	5.04
16	Modified Alg.	Dynamic	71	69.23	70.09	12.15	9.20	64.02	5.52
17	Modified Alg.	None	87	79.49	71.79	24.66	20.67	105.61	5.37
18	Modified Alg.	Fixed	87	80.34	78.63	18.06	14.62	75.88	4.86
19	Modified Alg.	Dynamic	87	67.52	74.36	12.72	9.47	62.25	5.53
20 (Whole Image)	Original Alg.	None	29	68.38	64.96	71.35	93.60	499.54	4.20

4. CONCLUSIONS

The goal of this research is to develop a computer-assisted tool that provides digitized histology image CIN discrimination capability in a "reasonable" time for potential application as a diagnostic aide in a clinical setting. In previous research [7], image analysis techniques were developed for CIN discrimination of digitized histology images. Examining the computation times for the steps in the analysis process for the experimental data set, the nuclei detection image processing algorithms were determined to be the most computationally intensive steps. In this study, we investigated approaches to improve the computational efficiency for different facets of the nuclei detection process to reduce the execution time for extracting features from the digitized histology images, including image resizing to reduce image resolution, adjusting the window size used in the clustering algorithm for mask generation in detecting nuclei, and modifying the background inversion approach for the nuclei detection preprocessing operations. In addition, image analysis operations from the implementation in [7] were vectorized to facilitate execution time improvement that did not change the feature calculations and CIN classifications. We investigated the impact of the three algorithm modifications on the timing and CIN classification accuracy of 117 digitized histology images. Experimental results from Table 1 showed that all proposed methods reduced feature extraction time. Notably, in Experiment 15, we obtained accuracy better than the baseline, 80.34%, with a simultaneous improvement in execution time with average of 17.56 seconds. Future research may include the use of other classifiers for fusion of the individual vertical segment classification results to form the final diagnosis.

acknowledge the medical expertise and collaboration of Dr. Mark Schiffman and Dr. Nicolas Wentzensen, both of the National Cancer Institute's Division of Cancer Epidemiology and Genetics (DCEG).

REFERENCES

[1] Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D. M., Forman, D., and Bray, F., 'Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012,' International Journal of Cancer, 2015, **136**(5), pp. E359–E386, ISSN 00207136, doi:10.1002/ijc.29210.

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

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

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

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

[6] Egner, J. R., 'AJCC Cancer Staging Manual,' JAMA, 2010, **304**(15), p. 1726, ISSN 0098-7484, doi:10.1001/jama.2010.1525.

[7] Guo, P., Banerjee, K., Stanley, R., Long, R., Antani, S., Thoma, G., Zuna, R., Frazier, S. R. S., Moss, R. H., and Stoecker, W. V., 'Nuclei-Based Features for Uterine Cervical Cancer Histology Image Analysis with Fusion-based Classification,' IEEE Journal of Biomedical and Health Informatics, 2015, **20**(6), pp. 1595–1607, ISSN 2168-2208, doi:10.1109/JBHI.2015.2483318.

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

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

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

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

[12] Park, S. C., Tan, J., Wang, X., Lederman, D., Leader, J. K., Kim, S. H., and Zheng, B., 'Computer-aided detection of early interstitial lung diseases using low-dose CT images,' Physics in medicine and biology, 2011, **56**(4), pp. 1139–53, ISSN 1361-6560, doi:10.1088/0031-9155/56/4/016.

[13] MathWorks, 'Matlab Profiler,' 2012.

[14] Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P., *Molecular Biology of the Cell*, Garland Science, 4th edition, 2002.

III. CLASSIFICATION OF UTERINE CERVICAL CANCER DIGITAL HISTOLOGY IMAGES USING A HYBRID DEEP LEARNING AND HANDCRAFTED FEATURES APPROACH

Haidar A. Almubarak[1], Peng Guo[1], R. Joe Stanley[1], Rodney Long[2] , Sameer Antani[2],

George Thoma[2], Rosemary Zuna[3], Shelliane R. Frazier[4]

[1]Department of Electrical & Computer Engineering

Missouri University of Science and Technology Rolla, Missouri 65409–0050

Tel: 573–341–6896

[2]Lister Hill National Center for Biomedical Communications for National Library of

Medicine, National Institutes of Health, DHHS in Bethesda, MD, USA

[3] Department of Pathology for the University of Oklahoma Health Sciences Center,

Oklahoma City,OK, USA

[4]Surgical Pathology Department for the University of Missouri Hospitals and Clinics,

Columbia, MO, USA

Email: h.almubarak@ieee.org, {pgp49,stanleyj}@mst.edu,

{rlong,antani,gthoma}@mail.nih.gov

ABSTRACT

Cervical cancer is the second most common cancer affecting women worldwide. Cervical cancer is curable if diagnosed early. Routinely, expert pathologists visually examine histology slides for assessing cervix tissue abnormalities. This research introduced an automated localized, fusion-based algorithm to classify squamous epithelium into normal, CIN1, CIN2, and CIN3 grades of cervical intraepithelial neoplasia (CIN). The approach was developed based on ten vertical partitions (segments) of the entire epithelium region obtained using machine vision algorithms and rectangular patches from the vertical segments

generated using a sliding window in deep learning preprocessing. The rectangular patch features were created based on both RGB and LAB color spaces to classify the segment and the result was fused to classify the whole epithelium. This research extends the previous research by dividing each of the ten segments into top, middle, and bottom sections and uses a convolutional neural network to classify these three sections. The result was then fused to classify the segments and the whole epithelium. The experimental data consisted of 83 images. With the proposed hybrid method, the CIN exact class classification accuracy was increased by 8.4%, and achieved test accuracy results 80.72% compared to 72.29% using the benchmark feature-based method for the same dataset.

Keywords: Cervical Cancer, Convolutional Neural Networks, Data Fusion, Image Classification, Clinical Decision Support Systems, Deep Learning, Feature Extraction

1. INTRODUCTION

Cervical cancer is the second leading cause of cancer death in women aged 20 to 39 years and was responsible for 8.8 million deaths in 2015 [1]. Screening or diagnosis for cervical cancer and its precursor lesions is carried out using a Papanicolaou (Pap) test. Biopsied cervical tissue histology slides are interpreted visually by the expert pathologist to give a definitive evaluation [2]. In the early diagnosis of cervical cancer, pathologists seek to detect cervical intraepithelial neoplasia (CIN), which is a pre-malignant condition for cervical cancer. A cervical biopsy is classified as normal (no CIN) or as CIN1 (mild dysplasia), CIN2 (moderate dysplasia), or CIN3 (severe dysplasia) by identifying the atypical cells in the epithelium by the visual inspection of histology slides [3]. Figure 1 shows an example of different CIN grades. Delayed maturation with an increase in immature atypical cells from bottom to top of the epithelium has been observed as CIN increases in severity [4]. Computer-assisted CIN diagnoses have been studied and developed pre-

viously in [5, 6, 7, 8, 9, 10]. In these studies handcrafted features were extracted using time-consuming image processing and machine learning algorithms, which may not provide the best accuracy.
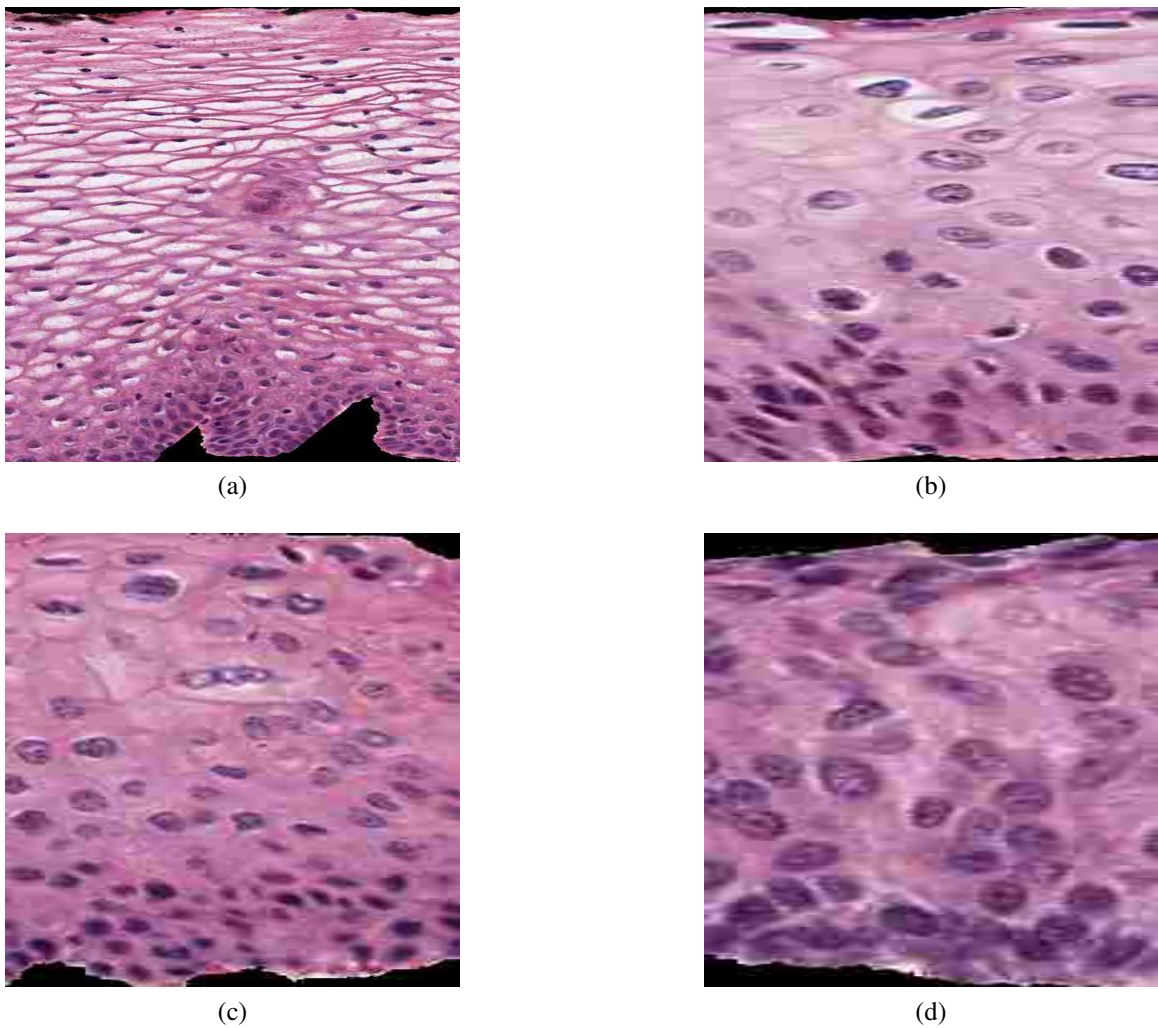
Figure 1. Example of CIN Grades: (a) Normal, (b) CIN1, (c) CIN2, (d) CIN3.

Convolutional neural networks (ConvNets) comprise a powerful technique used in many image analysis and classification problem domains with very large-scale datasets like ImageNet[11], face recognition [12], and breast cancer mitosis detection [13]. ConvNet is a feature-free technique that utilizes filters and convolves them with the images to extract the features. The filters are updated and tuned during the training process. In previous research, our research group used a localized fusion-based approach for CIN grade classification

[6]. This localized approach partitioned an epithelium image into ten vertical segments (partitions), where 27 features were extracted from each vertical segment [6, 7]. The features were used to classify each segment into one of the CIN grades. This research extends the localized fusion-based approach by further subdividing each segment into three sections: top, middle, and bottom. This division exploits the progression in cellular abnormality from the bottom (basal membrane) to the top of the epithelium by analyzing the three sections separately, then fusing the results to improve classifier performance. In this research, as many as 12 ConvNet output parameters were used as a feature set, while working together with the previously developed 27 image-based features from [6, 7]. All the features were processed by classifiers such as SVM, LDA, and Random Forest, and trained based on the ten vertical segments. After classifying the segments, the decision for the whole image class was determined by voting among the ten segments..

2. METHODOLOGY

In this paper, 83 full-color digitized histology images were used with Hematoxylin and Eosinophil (H&E) stain preparations of tissue sections of normal cervical tissue, and the three grades of cervical carcinoma in situ, which extends the data set in the previous research from [6]. The epithelium region of each image was extracted using a manually drawn border, which was confirmed by pathologists. Each epithelium region in a given image was vertically partitioned into ten sections (vertical segments) and processed to generate hand-crafted features using image processing and machine vision techniques as well as deep learning features. Based on the expert pathologist (RZ) ground truth labels, machine-learning algorithms were trained using the features extracted from the vertical segments as input. The steps for processing an epithelium image for CIN classification are follows:

1. The whole image was divided into ten segments.

2. The hand-crafted features were extracted from each segment (the 27 features).

3. Each vertical segment was divided into three sections: top, middle, bottom.

4. 64x64 RGB and LAB color space patches were extracted from each section using a sliding window with 50% overlap.

5. For each color space, three ConvNets were trained using the 64x64 patches, one for each section (top, middle, bottom).

6. The training and testing patches were classified with the trained model to get classification probability.

7. The three sections' class probabilities were obtained based on the patches extracted from each section. This formed a 12-features vector for each color space.

8. A new set of feature vectors were formed by concatenating the color space features and the 27 handcrafted features from each segment.

9. A classifier was trained to classify each segment for each feature vector and the segment class with probability was obtained.

10. A voting scheme was used to classify each image based on the maximum class occurrence among the segments.

11. A new 40-feature vector was constructed from the class probabilities of the ten segments.

12. A new classifier was trained on the 40 features and used to classify the test images.

The remainder of this section presents each step in detail.

2.1. The Handcrafted 27 Features. The segmentation of the 83 epithelium images used the work-flow employed in [6, 7], with the following five-step approach:

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

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

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

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

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

The following sections present each step in detail.

2.2. Medial Axis Detection and Segments Creation. The method for computing the medial axis, which is based on the distance transform, is presented in detail in [6, 7].The resulting medial axis was partitioned into ten vertical segments of approximately equal length. Perpendicular line slopes were estimated at the mid-points of each segment, and vertical lines were projected at the end points of each segment to generate ten vertical segments for analysis. The epithelium image was partitioned into ten vertical segments to facilitate localized diagnostic classification on sub-regions within the epithelium.
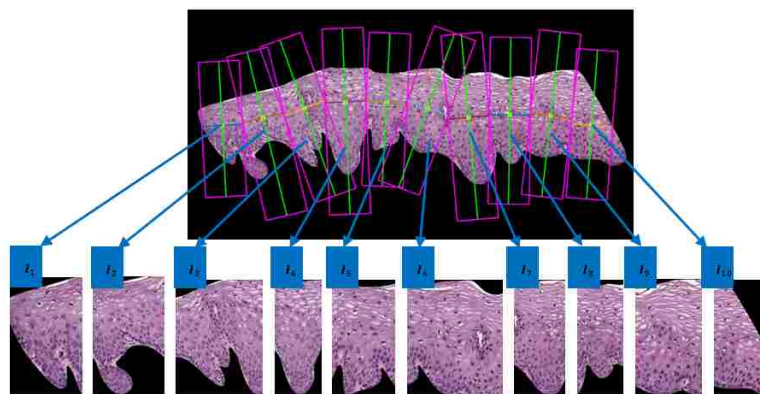
Figure 2. Epithelium Image Example with Vertical Segment Images (I_1 ,I_2, ,...., I_{10}) Determined From Bounding Boxes After Dividing the Medial Axis Into Ten Line Segment Approximations After Medial Axis Computation.

2.3. Feature Extraction. Features were obtained for each of the ten vertical segments of the whole image I_1, I_2, I_3, . . . , I_{10}. All the segments of one whole image were feature extracted in a sequence, in the order from left to right (see Figure. 2). The five different types of features that were extracted include: 1) texture features (F1-F10); 2) cellularity features (F11-F13); 3) nuclei features (F14, F15); 4) acellular (light area) features (F16-F22); 4) combination features (F23, F24); and 5) advanced layer-by-layer triangle features (F25- F27) [6, 7]. Table 1 is presented to provide a brief introduction, showing the feature label and brief description for each feature obtained.

2.3.1. Texture and cellular features. The texture and color features were used in previous work and was described in [6, 7]. The use of color in histopathological image analysis was also described in [8] and [9]. For texture features, both first-order structural measures and second-order statistical methods based on the gray-level cooccurrence matrix (GLCM) [5, 14] were used. The texture features include contrast (F1), energy (F2), correlation (F3), and uniformity (F4) of the segmented epithelial region, combined with the same statistics (contrast, energy, and correlation) obtained from the GLCM of the segment (F5-F10, see Table 1).
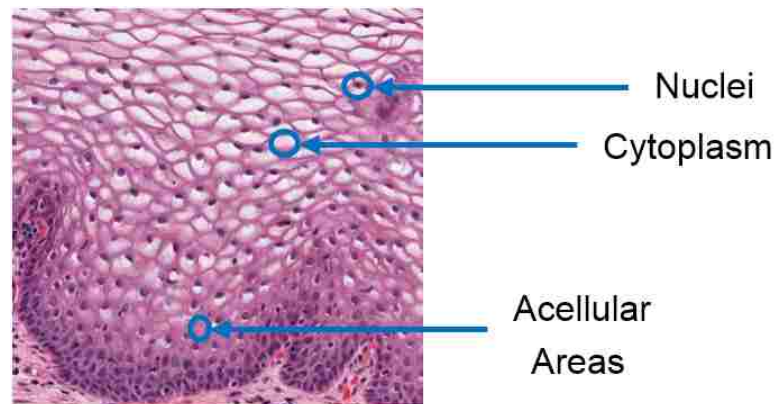
Figure 3. Sample Shading Representatives Within Epithelium Image Used for Determining Cellular Features.

Table 1. Feature Description

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

As shown in Figure 3 of normal cervical histology, the luminance images with three different regions of intensity (light, medium and dark) are marked within each region of interest. The light areas correspond to acellular areas, the medium areas to cytoplasm, and the dark areas to nuclei.

The cluster centers were located from the luminance image using K-means clustering [6, 7] for three different regions (K = 3) denoted as $clustLight$, $clustMedium$, and $clustDark$ for the light, medium, and dark cluster centers, respectively. Then, the ratios were calculated based on equations (1)-(3) [7].

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

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

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

where Acellular ratio (F11), Cytoplasm ratio (F12), and Nuclei ratio (F13) represent the cellular features in Table 1, and $numLight$, $numMedium$, and $numDark$ represent the number of pixels that were assigned to the clusters of light, medium, and dark, respectively.

2.3.2. Nuclei features. The dark area features represent nuclei in epithelial cells, varying in shape and size. When the CIN level increases from normal, both the number and size of nuclei tend to increase correspondingly. This relationship between the nuclei characteristics and CIN levels motivates the development of nuclei detection and nuclei feature extraction. Specifically, the following procedures were carried out [6, 7]:

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

Step 2. Perform morphological image closing [6, 7] on the nuclei mask image.

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

Step 4. Apply morphological image opening with a circular structuring element of radius 4 on the image from Step 3

Step 5. Eliminate small area noise objects (nonnuclei objects) within the epithelium region obtained in Step 4.

2.3.3. Acellular features. Extracting features in the light area regions was challenging due to the color and intensity variations in the epithelium images. The research team evaluated each of the L*, a*, and b* channels of CIELAB color space for characterizing the light areas and determined empirically that L* provides the best visual results. The following outlines the methods that were used for light area segmentation.

- Step 1: Convert the original image from RGB color space to L*a*b* color space.

- Step 2: Perform adaptive histogram equalization on the image obtained from step 1 using MATLAB's *adapthisteq*, which operates on small regions (tiles) so that the histogram of the output region matches a specified histogram and combines neighboring tiles using bilinear interpolation to eliminate artificially induced boundaries.

- Step 3: After the image has been contrast adjusted, the image is *binarized* by thresholding with the grayscale threshold 0.6 on the range 0 to 1. This step is intended to eliminate the dark nuclei regions and to retain the lighter nuclei and epithelium along with the light areas.

- Step 4: Segment the light areas using the K-means algorithm based on [7], with K equal to 4. A light area clustering example is given in [6].

- Step 5: Remove all objects having an area less than 100 pixels from the image with implementation of the MATLAB function *regionprops*. A morphological closing is performed with a disk structure element of radius 2.

2.3.4. Combination features. After both the nuclei features and the acellular features were extracted, the combined features were calculated to capture the relative change in nuclei numbers as CIN grade either increases or decreases. These features were the ratio of the acellular number to the nuclei number (F23), and the ratio of the acellular area to the total nuclei area (F24). The equations for calculating the combination features are presented as follows:

$$Ratio\ acellular\ number\ to\ nuclei\ number = \frac{Light\ Number}{Nuclei\ Number} \qquad (4)$$

$$Ratio\ acellular\ area\ to\ nuclei\ area = \frac{Light\ Area}{Nuclei\ Area} \qquad (5)$$

where $Light Number$ and $Nuclei Number$ refer to the total number of light area and nuclei, respectively. (see Fig. 3).

2.3.5. Triangle features. In previous research, triangle features have been investigated based on the Delaunay triangle transformation (DT) [6, 7] to detect nuclei for use as vertices. After locating the vertices for DT, the DT algorithm iteratively selects point triples which become vertices of each new triangle created. Delaunay triangulation exhibits the property that no point (nuclei centroid) lies within the circles that are formed by joining the vertices of the triangles [5]. All the triangles in the three layers were unique and there was no interaction or overlap within the triangles. The features were obtained based on the triangles formed in three different layers from top to bottom, including the number of triangles in the top layer (F25), the number of triangles in the middle layer (F26), and the number of triangles in the bottom layer (F27).

2.4. Extracting the Patches for the Deep Learning Features. This work extended the work in [15]. In the previous work, a 32x32 patch size with RGB color space was used. For the experiments in this paper, the patch size was increased to 64x64 pixels, and LAB color space was used as well as RGB. In [15], a non-overlapping window was used and only

patches falling entirely inside the epithelium were considered. In this study, an overlapping window of 50% was used, and patches with up to 10% of background (non-epithelium) were included.

Each of the ten segments was divided into three sections: top, middle, and bottom. The partitioning of the epithelium image into ten vertical segments helped in improve the CIN assessment through fusion of local sub-region classifications of the epithelium [6, 7].

The hypothesis for dividing each segment into three horizontal sections was to take advantage of the local differences in the sections among the CIN classes to improve the overall epithelium classification accuracy as the atypical cells change from bottom to top with the progression of CIN grade [4].

The epithelium region inside a histology image has an irregular shape and a nonuniform size, which cannot be used directly as input to the convolutional neural network. To fix this, there are two general approaches: the first is to resize all sections to a fixed size, and the second is to extract patches of a fixed size from each section. Resizing the entire section has the potential of losing some of the small, important details of the image. For this research, the second approach was used by extracting 64x64 pixel patches from each section. The patches were extracted using an overlapping sliding window with 50% overlap over the three sections of each segment. This served two purposes: the details of the histology slide were kept intact, and more samples were generated to train the deep learning network. Figure 4 shows an example of a segment being divided into three sections and the 64x64 patches being extracted from the bottom section. The low contrast squares in the 64x64 image indicate that they were not selected because they contain background area more than the allowed 10%. For illustration purposes the example shows the non-overlapping patches. In the experiments a 50% overlap extracted a new patch from each two adjoining patches that were selected with the non-overlapping window, and it might extract a new patch between a selected patch and a non-selected patch. Each patch was converted to LAB color space; both RGB and LAB patches were retained.

The Whole segment

Top Section

Middle Section

Bottom Section

64 by 64 patches from the segments' bottom section
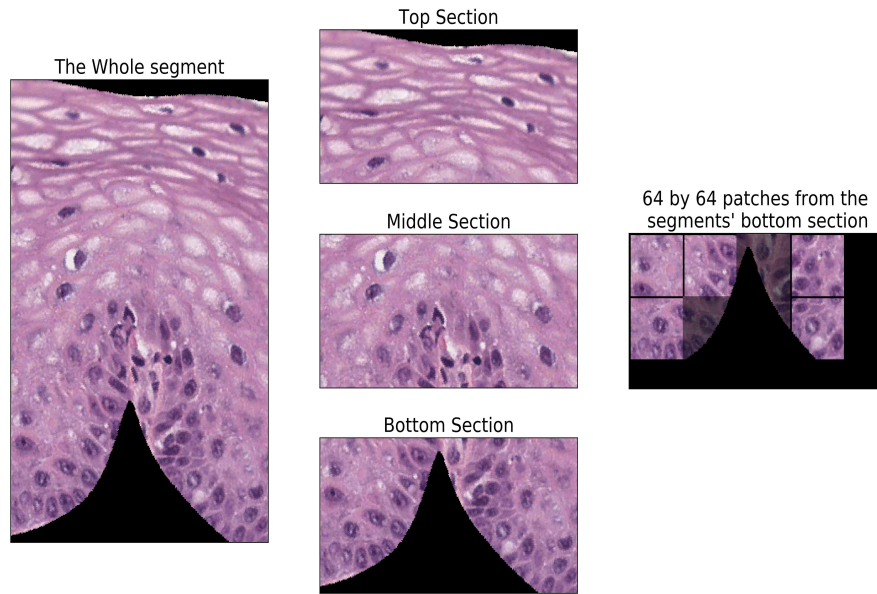
Figure 4. Example of Vertical Segment Being Divided into 3 Sections and Examples of 64x64 Patches Extracted

2.5. Training Convolutional Neural Network. The 64x64 patches from each section/color space were used to train a different ConvNet. There were the topNet, midNet, and botNet for each color space, which made a total of six ConvNets to train. Each network was trained using the patches that correspond to its section and color space of the segment. All networks have the same configuration as listed in Table 2. The input to the networks were the raw RGB and LAB values of the 64x64 patch that is followed by a series of convolution and max pooling layers with dropout layers in between. The convolution and max pooling layers were used to extract features from each patch by changing the convolution layers' filter weights. The last three layers were dense (fully connected) layers used to classify the image (patch) into one of four classes using features extracted from previous layers. The trained networks were then used to classify the patches from the test set. This architecture was chosen after experimenting with several architectures and different hyper-parameters; the selected architecture performed the best in the dataset.

Table 2. Convolution Neural Network Configuration

Layer type	Layer properties
Input	Size (3,64,64)
Conv2D	32 filters, size (3,3)
Conv2D	32 filters, size (3,3)
Max-poll	Size (2,2)
Dropout	P(0.25)
Conv2D	64 filters, size (3,3)
Max-poll	Size (2,2)
Dropout	P(0.25)
Conv2D	128 filters, size (3,3)
Max-poll	Size (2,2)
Dropout	P(0.25)
Flatten	convert last Conv output to a vector
Dense	500 Nodes
Dropout	P(0.25)
Dense	500 Nodes
Dense (Output)	4 Nodes

2.6. Segment Classification. The ConvNet classified each 64x64 patch into Normal, CIN1, CIN2, or CIN3. The segment sizes are not uniform, hence the number of patches differ between segments. To get a meaningful comparison/classification of the segments, the output of all segments needs to be put into a uniform feature vector. The feature vector was computed using the following formula:

$$class \; X_{-p}(y) = \frac{\textit{\# of patches classified as x}}{\textit{total number of patches in section Y}} \tag{6}$$

That is, the probability of section Y (top, middle, and bottom) being from class x (Normal, CIN1, CIN2, CIN3) calculated as the ratio of patches belonging to class x to the total number of patches in section Y. Using this formula, a feature vector of size 12 is created for each segment using the RGB color space and another 12-feature vector for LAB color space. Table 3 shows an example of this vector.

Table 3. Segment Classification Probability Distribution (12-feature vector)

Top Third				Middle Third				Bottom Third			
1	2	3	4	5	6	7	8	9	10	11	12
Normal_p	CIN1_p	CIN2_p	CIN3_p	Normal_p	CIN1_p	CIN2_p	CIN3_p	Normal_p	CIN1_p	CIN2_p	CIN3_p

In other words, the feature vector is the percentage of patches falling in each class (normal, CIN 1-3). Each section (top, middle, and bottom) of the segment will be classified independently using the ConvNet model trained for that section. Each segment will have three feature vectors, the first is a 12-feature vector similar to the one in Table 3 from the RGB color space, the second is another 12-feature vector from the LAB color space, and the last is the 27 features extracted based on the method described in Section 2.1. The three vectors are used to form a combination of features for each segment as follows

1. Deep learning features based on *RGB* color space only (12 features total)

2. Deep learning features based on *LAB* color space only (12 features total)

3. RGB + LAB (24 features)

4. Hand-crafted features from Section 2.1 (27 features total)

5. RGB + the 27 features (39 features total)

6. LAB + the 27 features (39 features total)

7. RGB + LAB + the 27 features (51 features total)

These vectors were used as a training input to other classification algorithms namely support vector machine (SVM), linear discriminant analysis (LDA), multilayer perceptron (MLP), logistic regression, and random forest (RF). These algorithms produce a classification probability value and a class for each segment.

2.7. Whole Epithelium Image Classification. The whole epithelium class is assigned using two approaches, the first approach uses voting among the ten segments classes, and the most frequent class is assigned to the whole epithelium. For a given feature set there are five classification scores, one from each classifier, and the final class from each classifier is used to give a vote for the whole epithelium image and compared to the individual classifiers. The second approach uses the four classification probabilities from Section 2.6 to construct a new feature vector from each classifier for each of the seven feature sets. The vector length is 40 (4 classes from 10 segments). This feature vector is used to train new classifiers (SVM, LDA, MLP, logistic, and RF) that are used to classify the whole epithelium images. The final result of all classifiers in a given feature set is also used in a voting scheme to classify the epithelium and is compared to the individual classifier output.

3. EXPERIMENTS PERFORMED AND RESULTS

The dataset used in these experiments consisted of 83 epithelium images. The data set contained 37 images classified as normal, 17 images classified as CIN1, 15 images classified as CIN2, and 14 images classified as CIN3. The images have been annotated by an expert pathologist. The ConvNets for RGB and LAB color spaces were trained using five-fold cross-validation. The data set was divided into 80% training and 20% testing sets, where the test sets were disjointed. The patches were extracted according to the method described in Section 2.4. The patch level exact class classification accuracy on the test images for RGB color space can be found in Table 4. Table 5 shows the average result when using the LAB color space. The LAB color space accuracy was on average less than RGB, which will result in lower segment accuracy.

Table 4. ConvNet 64x64 RGB Patches Test Set Average Classification Result

Fold	Top	Middle	Bottom
1	78.78%	60.88%	58.27%
2	55.94%	50.24%	40.37%
3	67.97%	65.50%	64.53%
4	82.82%	72.73%	64.94%
5	46.69%	42.15%	36.30%
Average	**66.44%**	**58.30%**	**52.88%**

Table 5. ConvNet 64x64 LAB Patches Test Set Average Classification Result

Fold	Top	Middle	Bottom
1	69.84%	63.54%	59.06%
2	49.13%	40.76%	29.61%
3	65.68%	60.34%	60.25%
4	82.47%	70.42%	62.39%
5	47.75%	39.76%	41.30%
Average	**62.97%**	**54.96%**	**50.52%**

The bottom section of the segment has a lower accuracy in general. Only folds 1 and 3 have accuracy comparable to the other sections of the segment. That is due to the similarity of the basal membrane nuclei distribution on the bottom section among the different classes; it is hard for the neural network to extract features that distinguish them.

Extracting the patches using the method described in Section 2.4 produced more than 90000 patches for network training and testing based on the five-fold cross-validation approach. The patch classification from the test set of images were fused using the method

described in Section 2.4 (Equation 6 in particular) to generate top, middle and bottom sections CIN classification probabilities for each vertical segment and to produce the 12-feature vector (Table 3) for each color space from each segment.

For direct comparison with the previous feature-based vertical segment fusion for the whole image CIN classification [6, 7], a leave-one-image-out approach was examined. This approach was used to accommodate for the limited number of epithelium images in the data set, particularly the limited number of epithelium images in the different CIN classes.

The seven feature sets (Deep Learning RGB 12 features, Deep Learning LAB 12 features, RGB+LAB combined 24 features, 27 handcrafted features, RGB + 27 combined, LAB + 27 combined, RGB + LAB + 27 combined) described in Section 2.6 were created for each segment and were used to train each of the classifiers (SVM, LDA, ... etc.). Table 6 shows the accuracy of segment classification for the different feature sets when using different classification algorithms.

Table 6. Vertical Segments Classification Accuracy Based on Leave-one-image-out

Features Set	SVM	LDA	MLP	Logistic	RF
RGB (deep learning features) only	63.93%	62.97%	60.92%	61.28%	61.76%
LAB (deep learning features) only	58.62%	56.33%	56.33%	54.16%	59.47%
27 (handcrafted) features only	66.95%	67.79%	62.97%	64.17%	62.85%
RGB+LAB combined	61.40%	62.00%	58.63%	61.88%	60.68%
RGB + 27 combined features	**71.29%**	69.36%	65.02%	68.88%	65.14%
LAB + 27 combined features	**71.29%**	69.24%	65.26%	66.71%	65.14%
RGB + LAB + 27 combined features	69.48%	69.72%	69.24%	69.00%	65.14%

The best accuracy was achieved using the SVM classifier with RGB + 27 features and LAB + 27 features (71.29%). Using the LDA classifier with the same features gave a close accuracy of 69.36% and 69.24%, respectively. Combining the RGB, LAB, and the 27 features gave accuracies in the same range when using SVM, LDA, and MLP classifiers.

The segment classifications from an image were used to vote for the whole epithelium class (first approach), and the results are shown in Table 7. In the voting scheme, the number of segments in a given class were counted and the whole image class is determined by the class that occurred the most. If there was a tie the higher class was assigned. For example, if there are five segments classified as CIN1 and five classified as CIN2, then the whole epithelium is assigned CIN2 class.

The underlined values correspond to the method used in [7]. This approach gave a classification accuracy of 69.87% and 72.29% using SVM and LDA classifiers, respectively. Tables 9 and 10 show the confusion matrices for these classifiers. The best classification accuracy was 80.72% when using the SVM classifier with the deep learning LAB combined with the 27 handcrafted features. Table 11 shows the confusion matrix for this classifier. Although the RGB + 27 features with the SVM classifier on the segment level scored the same as as the LAB + 27 features, the voting among the the segments showed that the LAB + 27 features had a 5% increase on the whole epithelium classification accuracy. This is because the voting scheme needs the majority (e.g. 6 out of 10) of segments to be classified correctly to give a correct class for the image. If some images have all of their 10 segments correctly classified and the other images have less than 6 out 10 correctly classified, then the segment level accuracy might be high but is not necessary reflected on the whole image classification accuracy. Combining the deep learning RGB, LAB and the 27 features and using SVM classifier gave the same accuracy as using RGB and the 27 features only. Table 12 shows the confusion matrix for this classifier/feature combination. Using the LDA classifier with all deep learning and handcrafted features gave a higher accuracy of 77.11%. Table 13 shows the confusion matrix when using LDA classifier with the deep learning RGB + LAB and the 27 features as inputs.

Table 7. Whole Epithelium Classification Using the Segment Class Voting Scheme

Features Set	SVM	LDA	MLP	Logistic	RF
RGB (deep learning features) only	69.88%	63.86%	60.24%	63.86%	69.88%
LAB (deep learning features) only	62.65%	59.04%	59.04%	57.83%	60.24%
27 (handcrafted) features only	_69.87%_	_72.29%_	66.27%	63.86%	63.86%
RGB+LAB combined	62.65%	62.65%	60.24%	69.88%	60.24%
RGB + 27 combined features	75.90%	77.11%	67.47%	72.29%	67.47%
LAB + 27 combined features	**80.72%**	77.11%	69.88%	69.88%	62.65%
RGB + LAB + 27 combined features	75.90%	77.11%	73.49%	73.49%	67.47%
Voting among classifier	73.49%	75.90%	66.27%	71.08%	66.27%

Table 8. Whole Epithelium Classification Using the Segment Class Voting Scheme with Top and Middle Sections Only

Features Set	SVM	LDA	MLP	Logistic	RF
RGB (deep learning features) only	63.86%	61.45%	57.83%	62.65%	68.67%
LAB (deep learning features) only	55.42%	59.04%	61.45%	51.81%	59.04%
27 (handcrafted) features only	69.88%	72.29%	59.04%	63.86%	63.86%
RGB+LAB combined	61.45%	67.47%	59.04%	61.45%	62.65%
RGB + 27 combined features	74.70%	74.70%	55.42%	68.67%	66.27%
LAB + 27 combined features	73.49%	72.29%	55.42%	67.47%	69.88%
RGB + LAB + 27 combined features	68.67%	72.29%	61.45%	67.47%	65.06%
Voting among classifier	72.29%	69.88%	57.83%	65.06%	63.86%

Despite the low accuracy of the bottom section of the segments, the bottom section of the vertical segments is important in the overall image classification. Table 8 shows the accuracy of the whole epithelium classification using only the top and middle sections' deep learning features of the segments to train the classifier.

Table 9. Whole Epithelium Confusion Matrix for the 27 Features Using SVM Classifier and Voting Scheme

Classified as ⟹				
True class ⟱	Normal	CIN1	CIN2	CIN3
Normal	33	4	0	0
CIN1	8	7	2	0
CIN2	0	3	8	4
CIN3	1	0	3	10

Table 10. Whole Epithelium Confusion Matrix for the 27 Features Using LDA Classifier and Voting Scheme

Classified as ⟹				
True class ⟱	Normal	CIN1	CIN2	CIN3
Normal	33	3	0	1
CIN1	7	6	4	0
CIN2	0	3	11	1
CIN3	0	0	4	10

Table 11. Whole Epithelium Confusion Matrix for LAB and the 27 Features Using SVM Classifier and Voting Scheme

Classified as ⟹				
True class ⟱	Normal	CIN1	CIN2	CIN3
Normal	34	3	0	0
CIN1	5	10	2	0
CIN2	0	1	11	3
CIN3	0	0	2	12

Table 12. Whole Epithelium Confusion Matrix for RGB,LAB, and the 27 Features Using SVM Classifier and Voting Scheme

Classified as ⇒				
True class ⇓	Normal	CIN1	CIN2	CIN3
Normal	33	4	0	0
CIN1	9	6	2	0
CIN2	0	1	12	2
CIN3	1	0	1	12

Table 13. Whole epithelium Confusion matrix for RGB, LAB, and the 27 features using LDA classifier and voting scheme

Classified as ⇒				
True class ⇓	Normal	CIN1	CIN2	CIN3
Normal	32	5	0	0
CIN1	7	9	1	0
CIN2	0	3	11	1
CIN3	0	0	2	12

In the second approach, the segment class probabilities (the 40 features) from each classifier/feature set were used as inputs to another classifier. Each classifier(SVM, LDA, MLP, ...) output is used to train new SVM, LDA, MLP, Logistic, and Random Forest classifiers. For example the features generated using the SVM classifier on the segments level are used to train the new SVM, LDA, MLP, Logistic, and Random forest classifiers for the whole epithelium classification. Also, the features generated using the LDA classifier on the segments level were used to train the new SVM, LDA, MLP, Logistic, and Random forest classifiers for the whole epithelium classification, this was repeated for all segments' level classifiers. Table 14 shows the result of using different classifiers for the 40-feature vector extracted by using SVM classifier for segment level classification. The highest accuracy

was 78.31%. This result was obtained when the RGB+27 features were used on an SVM classifier to classify the segments, and the probability distribution result from that was used to train another SVM classifier.

Table 15 shows the confusion matrix for the SVM classifier. This classifier was able to classify CIN2 epithelium images better than all classifier/feature combinations by correctly classifying 13 out of 15 images.

Table 14. Different Whole Epithelium Classification Accuracies Using the 40 Features When Using SVM for Segment Classification

Features Set	SVM	LDA	MLP	Logistic	RF
RGB (deep learning features) only	68.68%	60.24%	63.86%	62.65%	61.45%
RGB (deep learning features) only	66.27%	59.04%	57.83%	57.83	57.83%
27 (handcrafted) features only	67.47	61.45%	66.27%	71.08%	75.90%
RGB+LAB combined	63.86%	65.06%	56.63%	61.45%	59.04%
RGB + 27 combined features	**78.31%**	71.08%	71.08%	74.70%	75.90%
LAB + 27 combined features	72.29%	69.88%	68.68%	72.23%	69.88%
RGB + LAB + combined features	72.23%	65.06%	69.88%	72.29%	69.88%
Voting among classifier	74.70%	72.29%	72.29%	71.08%	72.29%

Table 15. Whole Epithelium Confusion Matrix Using SVM to Classify Segments Based on RGB and the 27 Features and Using Another SVM on the Result to Classify the Whole Image

Classified as ⇒				
True class ⇓	Normal	CIN1	CIN2	CIN3
Normal	31	6	0	0
CIN1	7	10	0	0
CIN2	1	0	13	1
CIN3	1	0	2	11

The rest of the classifier/feature combinations did not produce any improved result, but there are some interesting observations. Some of the classifier/feature combinations were able to correctly classify all the normal epitheliums but at the expense of misclassifying all the CIN1 epitheliums. Table 16 shows an example of this when using LDA classifier to classify the segments with RGB+LAB as input features then using SVM to classify the whole image.

Table 16. Whole Epithelium Confusion Matrix Using LDA to Classify Segments Based on RGB+LAB Features and Using SVM on the Result to Classify the Whole Image

Classified as ⇒				
True class ⇓	Normal	CIN1	CIN2	CIN3
Normal	37	0	0	0
CIN1	15	0	2	0
CIN2	4	0	8	3
CIN3	0	0	4	10

When using an LDA segment classifier with RGB+LAB+27 features as inputs and then using and LDA classifier to classify the whole epithelium image, all the CIN3 epithelium images were identified correctly. Table 17 shows the confusion matrix for this classifier/feature combination.

Table 17. Whole Epithelium Confusion Matrix Using LDA to Classify Segments Based on RGB + LAB + 27 features and Using LDA on the Result to Classify the Whole Image

Classified as ⇒				
True class ⇓	Normal	CIN1	CIN2	CIN3
Normal	31	6	0	0
CIN1	7	8	2	0
CIN2	1	1	10	3
CIN3	0	0	0	14

It also found that combining the deep learning features with the handcrafted features improved the detection of CIN1 epithelium in most of the classifier combinations. As an example, by using logistic regression to classify the segments and using the multilayer perceptron (MLP) to classify the whole epithelium with RGB features, only one CIN1 epithelium was correctly classified (Table 18). When using LAB features none of the CIN1 was correctly classified (Table 19), and with the 27 handcrafted features only four were classified correctly (Table 20). When the 27 features were combined with RGB (Table 21), the accuracy of detecting CIN1 increased and nine images were correctly classified. When combining the 27 features with LAB (Table 22), nine CIN1 images were correctly classified.The overall accuracy increased in both cases. The highest CIN1 classification accuracy was obtained when using Random Forest classifier on the LAB+27 features for both segment classification and image classification. Table 23 shows the confusion matrix for this classifier. The overall accuracy for this classifier was 77.10% which is close to the best accuracy discussed earlier. This classifier was the best in separating the normal form CIN with an accuracy of 92.77%. If we consider both exact class classification and separation of CIN from normal epithelium as a criteria for choosing the classifier, then using the random forest for both segment and image classification with the LAB + 27 feature set will be the optimal choice.

Table 18. Whole Epithelium Confusion Matrix Using Logistic Regression to Classify Segments Based on RGB Features and Using MLP on the Result to Classify the Whole Image

Classified as ⇒				
True class ⇓	Normal	CIN1	CIN2	CIN3
Normal	35	2	0	0
CIN1	14	1	1	1
CIN2	4	0	7	4
CIN3	0	0	4	10

Table 19. Whole Epithelium Confusion Matrix Using Logistic Regression to Classify Segments Based on LAB Features and Using MLP on the Result to Classify the Whole Image

Classified as ⇒				
True class ⇓	Normal	CIN1	CIN2	CIN3
Normal	35	2	0	0
CIN1	15	0	1	1
CIN2	4	0	2	9
CIN3	0	0	5	9

Table 20. Whole Epithelium Confusion Matrix Using Logistic Regression to Classify Segments Based on The 27 Features and Using MLP on the Result to Classify the Whole Image

Classified as ⇒				
True class ⇓	Normal	CIN1	CIN2	CIN3
Normal	31	6	0	0
CIN1	9	4	4	0
CIN2	0	3	8	4
CIN3	1	0	6	7

Table 21. Whole Epithelium Confusion Matrix Using Logistic Regression to Classify Segments Based on RGB+ 27 Features and Using MLP on the Result to Classify the Whole Image

Classified as ⇒				
True class ⇓	Normal	CIN1	CIN2	CIN3
Normal	32	5	0	0
CIN1	6	9	2	0
CIN2	0	3	10	2
CIN3	0	0	5	9

Table 22. Whole Epithelium Confusion Matrix Using Logistic Regression to Classify Segments Based on LAB + 27 Features and Using MLP on the Result to Classify the Whole Image

Classified as ⇒				
True class ⇓	Normal	CIN1	CIN2	CIN3
Normal	34	2	1	0
CIN1	7	9	1	0
CIN2	0	2	9	4
CIN3	0	0	7	7

Table 23. Whole Epithelium Confusion Matrix Using Random Forest to Classify the Segments and the Whole Epithelium Based on LAB + 27 Features

Classified as ⇒				
True class ⇓	Normal	CIN1	CIN2	CIN3
Normal	35	2	0	0
CIN1	4	12	1	0
CIN2	0	1	8	6
CIN3	0	0	5	9

In all classifier/feature combinations, CIN1 epithelium images have low detection accuracy. A possible explanation is that CIN1 looks visually similar to the normal epithelium, which makes it hard to detect. Figure 5 shows an example of a misclassified CIN1 and a normal epithelium.

It is observed that misclassification of CIN1 as normal is a low-cost error, since the clinical path for CIN1 comprises further clinical observation; in contrast the clinical path for higher CIN grades entails treatment. Future work could include different weights for the different CIN grade errors based on the different clinical outcomes.

(a)

(b)

Figure 5. Example of a Normal Epithelium (a) and a CIN1 Misclassified as Normal (b)

4. CONCLUSION

In this paper, a deep learning and handcrafted feature fusion approach using additional fusion for localized features to classify cervical tissue histology images was developed. From the experimental results, it was shown that combining the two feature approaches increases the overall classification accuracy. Using the deep learning approach alone achieved the same classification accuracy as the handcrafted features; combining them resulted in better classification. Although the deep leaning paradigm shows a promise in many image classification tasks, the handcrafted feature paradigm still has its place in image analysis and classification

ACKNOWLEDGMENTS

REFERENCES

[1] Siegel, R. L., Miller, K. D., and Jemal, A., 'Cancer statistics, 2017,' CA: A Cancer Journal for Clinicians, 2017, **67**(1), pp. 7–30, ISSN 00079235, doi:10.3322/caac.21387.

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

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

[4] Egner, J. R., 'AJCC Cancer Staging Manual,' JAMA, 2010, **304**(15), p. 1726, ISSN 0098-7484, doi:10.1001/jama.2010.1525.

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

[6] Guo, P., Banerjee, K., Stanley, R., Long, R., Antani, S., Thoma, G., Zuna, R., Frazier, S. R. S., Moss, R. H., and Stoecker, W. V., 'Nuclei-Based Features for Uterine Cervical Cancer Histology Image Analysis with Fusion-based Classification,' IEEE Journal of Biomedical and Health Informatics, 2015, **20**(6), pp. 1595–1607, ISSN 2168-2208, doi:10.1109/JBHI.2015.2483318.

[7] Guo, P., Almubarak, H., Banerjee, K., Stanley, R. J., Long, R., Antani, S., Thoma, G., Zuna, R., Frazier, S. R., Moss, R. H., and Stoecker, W. V., 'Enhancements in localized classification for uterine cervical cancer digital histology image assessment,' Journal of Pathology Informatics, 2016, **7**(1), p. 51, ISSN 2153-3539, doi:10.4103/2153-3539. 197193.

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

[9] Guillaud, M., Adler-Storthz, K., Malpica, A., Staerkel, G., Matisic, J., Van Niekirk, D., Cox, D., Poulin, N., Follen, M., and MacAulay, C., 'Subvisual chromatin changes in cervical epithelium measured by texture image analysis and correlated with HPV,' Gynecologic Oncology, 2005, **99**(3 SUPPL.), pp. 16–23, ISSN 0090-8258, doi:10. 1016/j.ygyno.2005.07.037.

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

[11] Krizhevsky, A., Sutskever, I., and Hinton, G., 'ImageNet Classification with Deep Convolutional Neural Networks,' in 'Advances in Neural Information and Processing Systems (NIPS),' volume 25, pp. 1097–1105, 2012.

[12] Parkhi, O. M., Vedaldi, A., and Zisserman, A., 'Deep Face Recognition,' in 'Procedings of the British Machine Vision Conference 2015,' Section 3, ISBN 1-901725-53-7, 2015 pp. 1–41, doi:10.5244/C.29.41.

[13] Ciresan, D. C., Giusti, A., Gambardella, L. M., and Schmidhuber, J., 'Mitosis Detection in Breast Cancer Histology Images using Deep Neural Networks,' Medical Image Computing and Computer-Assisted Intervention (MICCAI 2013), 2013, pp. 411–418.

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

[15] Almubarak, H., Stanley, R. J., Long, R., Antani, S., Thoma, G., Zuna, R., and Frazier, S., 'Convolutional Neural Network Based Localized Classification of Uterine Cervical Cancer Digital Histology Images,' in 'Complex Adaptive Systems 2017 (Complex Adaptive Systems 2017),' volume 114, ISSN 18770509, 2017 pp. 281–287, doi:10.1016/j.procs.2017.09.044.

SECTION

2. SUMMARY AND CONCLUSIONS

In this dissertation localized features from medical images have been explored. The computational complexity of feature extraction and analysis was analyzed and bottlenecks were identified and mitigated while maintaining classification accuracy. A new approach combining deep learning and handcrafted features was developed for medical image analysis and classification. Localized features have been applied to dermoscopy images to classify them into benign and malignant. It was also applied to histology images of cervical tissue to detect a pre-malignant phase known as cervical intra-epithelial neoplasia.

It was demonstrated through experimental results that extracting localized features and fusing them to classify the whole image improved the accuracy. Experimental results also show that combining deep learning features gave better results when compared a to single paradigm.

REFERENCES

[1] Paredes, R., Pérez, J. C., Juan, A., and Vidal, E., 'Local Representations and a Direct Voting Scheme for Face Recognition,' in 'Proc Workshop on Pattern Rec and Inf Systems,' 2001 pp. 71–79, doi:10.1.1.16.3558.

[2] Fronthaler, H., Kollreider, K., and Bigun, J., 'Local Features for Enhancement and Minutiae Extraction in Fingerprints,' IEEE Transactions on Image Processing, 2008, **17**(3), pp. 354–363, ISSN 1057-7149, doi:10.1109/TIP.2007.916155.

[3] Altamura, D., Menzies, S. W., Argenziano, G., Zalaudek, I., Soyer, H. P., Sera, F., Avramidis, M., DeAmbrosis, K., Fargnoli, M. C., and Peris, K., 'Dermatoscopy of basal cell carcinoma: Morphologic variability of global and local features and accuracy of diagnosis,' Journal of the American Academy of Dermatology, 2010, **62**(1), pp. 67–75, ISSN 01909622, doi:10.1016/j.jaad.2009.05.035.

[4] Rahmatullah, B., Papageorghiou, a. T., and Noble, J. a., 'Integration of local and global features for anatomical object detection in ultrasound,' in 'Miccai,' volume 15, pp. 402–409, Springer, Berlin, Heidelberg, ISBN 978-3-642-33453-5, 2012, doi: 10.1007/978-3-642-33454-2{_}50.

[5] Tamaki, T., Yoshimuta, J., Kawakami, M., Raytchev, B., Kaneda, K., Yoshida, S., Takemura, Y., Onji, K., Miyaki, R., and Tanaka, S., 'Computer-aided colorectal tumor classification in NBI endoscopy: Using local features,' Medical Image Analysis, 2013, **17**(1), pp. 78–100, ISSN 13618415, doi:10.1016/j.media.2012.08.003.

[6] Shyu, C. R., Brodley, C. E., Kak, A. C., Kosaka, A., Aisen, A., and Broderick, L., 'Local versus global features for content-based image retrieval,' Proceedings. IEEE Workshop on Content-Based Access of Image and Video Libraries (Cat. No.98EX173), 1998, pp. 30–34, doi:10.1109/IVL.1998.694482.

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

[8] Fuchs, M., Wagner, M., Wischmann, H.-A., Köhler, T., Theißen, A., Drenckhahn, R., and Buchner, H., 'Improving source reconstructions by combining bioelectric and biomagnetic data,' Electroencephalography and Clinical Neurophysiology, 1998, **107**(2), pp. 93–111, ISSN 00134694, doi:10.1016/S0013-4694(98)00046-7.

[9] Williams, M. B., Judy, P. G., Gunn, S., and Majewski, S., 'Dual-Modality Breast Tomosynthesis,' Radiology, 2010, **255**(1), pp. 191–198, ISSN 0033-8419, doi:10. 1148/radiol.09091160.

[10] Garg, A. X., Adhikari, N. K. J., McDonald, H., Rosas-Arellano, M. P., Devereaux, P. J., Beyene, J., Sam, J., and Haynes, R. B., 'Effects of Computerized Clinical Decision Support Systems on Practitioner Performance and Patient Outcomes,' JAMA, 2005, **293**(10), p. 1223, ISSN 0098-7484, doi:10.1001/jama.293.10.1223.

[11] Muller, H., Reihs, R., Posch, A. E., Kremer, A., Ulrich, D., and Zatloukal, K., 'Data Driven GUI Design and Visualization for a NGS Based Clinical Decision Support System,' in '2016 20th International Conference Information Visualisation (IV),' IEEE, ISBN 978-1-4673-8942-6, ISSN 10939547, 2016 pp. 355–360, doi: 10.1109/IV.2016.79.

[12] Miranda, E., Aryuni, M., and Irwansyah, E., 'A survey of medical image classification techniques,' 2016 International Conference on Information Management and Technology (ICIMTech), 2016, (November), pp. 56–61, doi:10.1109/ICIMTech.2016. 7930302.

[13] Wan, J., Wang, D., Hoi, S. C. H., Wu, P., Zhu, J., Zhang, Y., and Li, J., 'Deep Learning for Content-Based Image Retrieval: A Comprehensive Study,' Proceedings of the ACM International Conference on Multimedia - MM '14, 2014, pp. 157–166, doi:10.1145/2647868.2654948.

[14] Roth, H. R., Lu, L., Farag, A., Shin, H.-C., Liu, J., Turkbey, E. B., and Summers, R. M., 'DeepOrgan: Multi-level Deep Convolutional Networks for Automated Pancreas Segmentation,' in 'Medical Image Computing and Computer-Assisted Intervention - Miccai 2015, Pt I,' volume 9349, pp. 556–564, ISBN 978-3-319-24553-9; 978-3-319-24552-2, 2015, doi:10.1007/978-3-319-24553-9{_}68.

[15] Song, Y., Zhang, L., Chen, S., Ni, D., Li, B., Zhou, Y., Lei, B., and Wang, T., 'A deep learning based framework for accurate segmentation of cervical cytoplasm and nuclei,' in '2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society,' volume 2014, IEEE, ISBN 978-1-4244-7929-0, ISSN 1557170X, 2014 pp. 2903–2906, doi:10.1109/EMBC.2014.6944230.

[16] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., and Thrun, S., 'Dermatologist-level classification of skin cancer with deep neural networks,' Nature, 2017, **542**(7639), pp. 115–118, ISSN 0028-0836, doi:10.1038/nature21056.

[17] Codella, N., Nguyen, Q.-B., Pankanti, S., Gutman, D., Helba, B., Halpern, A., and Smith, J. R., 'Deep Learning Ensembles for Melanoma Recognition in Dermoscopy Images,' arXiv cs.CV, 2016, **10**(4), p. 04662.

[18] Ciresan, D. C., Giusti, A., Gambardella, L. M., and Schmidhuber, J., 'Mitosis Detection in Breast Cancer Histology Images using Deep Neural Networks,' Medical Image Computing and Computer-Assisted Intervention (MICCAI 2013), 2013, pp. 411–418.

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

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

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

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

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

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

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

[26] Bajaj, S., Marchetti, M. A., Navarrete-Dechent, C., Dusza, S. W., Kose, K., and Marghoob, A. A., 'The Role of Color and Morphologic Characteristics in Dermoscopic Diagnosis,' JAMA Dermatology, 2016, **152**(6), p. 676, ISSN 2168-6068, doi:10.1001/jamadermatol.2016.0270.

[27] Xu, R. and Wunsch, D. C., *Clustering*, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2008, ISBN 9780470276808, doi:10.1002/9780470382776.

[28] Aneja, D. and Rawat, T. K., 'Fuzzy Clustering Algorithms for Effective Medical Image Segmentation,' International Journal of Intelligent Systems and Applications, 2013, **5**(11), pp. 55–61, ISSN 2074904X, doi:10.5815/ijisa.2013.11.06.

[29] Qi, X., Wang, D., Rodero, I., Diaz-Montes, J., Gensure, R. H., Xing, F., Zhong, H., Goodell, L., Parashar, M., Foran, D. J., and Yang, L., 'Content-based histopathology image retrieval using CometCloud,' BMC Bioinformatics, 2014, **15**(1), p. 287, ISSN 1471-2105, doi:10.1186/1471-2105-15-287.

[30] Xu, R., Damelin, S., and Wunsch, D. C., 'Clustering of cancer tissues using diffusion maps and fuzzy ART with gene expression data,' 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), 2008, pp. 183–188, doi:10.1109/IJCNN.2008.4633787.

[31] Guo, P., Banerjee, K., Stanley, R., Long, R., Antani, S., Thoma, G., Zuna, R., Frazier, S. R. S., Moss, R. H., and Stoecker, W. V., 'Nuclei-Based Features for Uterine Cervical Cancer Histology Image Analysis with Fusion-based Classification,' IEEE Journal of Biomedical and Health Informatics, 2015, **20**(6), pp. 1595–1607, ISSN 2168-2208, doi:10.1109/JBHI.2015.2483318.

[32] Krizhevsky, A., Sutskever, I., and Hinton, G., 'ImageNet Classification with Deep Convolutional Neural Networks,' in 'Advances in Neural Information and Processing Systems (NIPS),' volume 25, pp. 1097–1105, 2012.

[33] Parkhi, O. M., Vedaldi, A., and Zisserman, A., 'Deep Face Recognition,' in 'Procedings of the British Machine Vision Conference 2015,' Section 3, ISBN 1-901725-53-7, 2015 pp. 1–41, doi:10.5244/C.29.41.

[34] Marghoob, A. A. and Scope, A., 'The complexity of diagnosing melanoma.' The Journal of Investigative Dermatology, 2009, **129**(1), pp. 11–3, ISSN 1523-1747, doi:10.1038/jid.2008.388.

[35] Stoecker, W. V. and Stolz, W., 'Dermoscopy and the Diagnostic Challenge of Amelanotic and Hypomelanotic Melanoma,' Archives of Dermatology, 2008, **144**(9), pp. 1207–10, ISSN 0003-987X, doi:10.1001/archderm.144.9.1207.

[36] Menzies, S. W., Moloney, F. J., Byth, K., Avramidis, M., Argenziano, G., Zalaudek, I., Braun, R. P., Malvehy, J., Puig, S., Rabinovitz, H. S., Oliviero, M., Cabo, H., Bono, R., Pizzichetta, M. A., Claeson, M., Gaffney, D. C., Soyer, H. P., Stanganelli, I., Scolyer, R. A., Guitera, P., Kelly, J., McCurdy, O., Llambrich, A., Marghoob, A. A., Zaballos, P., Kirchesch, H. M., Piccolo, D., Bowling, J., Thomas, L., Terstappen, K., Tanaka, M., Pellacani, G., Pagnanelli, G., Ghigliotti, G., Ortega, B. C., Crafter, G., Ortiz, A. M. P., Tromme, I., Karaarslan, I. K., Ozdemir, F., Tam, A., Landi, C., Norton, P., Kaçar, N., Rudnicka, L., Slowinska, M., Simionescu, O., Di Stefani, A., Coates, E., and Kreusch, J., 'Dermoscopic Evaluation of Nodular Melanoma,' JAMA Dermatology, 2013, **149**(6), p. 699, ISSN 2168-6068, doi:10.1001/jamadermatol.2013.2466.

[37] Soon, S. L., Solomon, A. R., Papadopoulos, D., Murray, D. R., McAlpine, B., and Washington, C. V., 'Acral lentiginous melanoma mimicking benign disease: The Emory experience,' Journal of the American Academy of Dermatology, 2003, **48**(2 SUPPL.), pp. 183–188, ISSN 01909622, doi:10.1067/mjd.2003.63.

[38] Kittler, H., Pehamberger, H., Wolff, K., Binder, M., and Levenstein, M., 'Diagnostic accuracy of dermoscopy.' The Lancet. Oncology, 2002, **3**(3), pp. 159–65, ISSN 1470-2045, doi:10.1016/S1470-2045(02)00679-4.

[39] Vestergaard, M. E., Macaskill, P., Holt, P. E., and Menzies, S. W., 'Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: A meta-analysis of studies performed in a clinical setting,' British Journal of Dermatology, 2008, **159**(3), pp. 669–676, ISSN 00070963, doi:10.1111/j.1365-2133.2008.08713.x.

[40] Ferris, L. K., Harkes, J. A., Gilbert, B., Winger, D. G., Golubets, K., Akilov, O., and Satyanarayanan, M., 'Computer-aided classification of melanocytic lesions using dermoscopic images,' Journal of the American Academy of Dermatology, 2015, **73**(5), pp. 769–776, ISSN 10976787, doi:10.1016/j.jaad.2015.07.028.

[41] Carrera, C., Marchetti, M. A., Dusza, S. W., Argenziano, G., Braun, R. P., Halpern, A. C., Jaimes, N., Kittler, H. J., Malvehy, J., Menzies, S. W., Pellacani, G., Puig, S., Rabinovitz, H. S., Scope, A., Soyer, H. P., Stolz, W., Hofmann-Wellenhof, R., Zalaudek, I., and Marghoob, A. A., 'Validity and Reliability of Dermoscopic Criteria Used to Differentiate Nevi From Melanoma,' JAMA Dermatology, 2016, **152**(7), p. 798, ISSN 2168-6068, doi:10.1001/jamadermatol.2016.0624.

[42] Argenziano, G., Puig, S., Zalaudek, I., Sera, F., Corona, R., Alsina, M., Barbato, F., Carrera, C., Ferrara, G., Guilabert, A., Massi, D., Moreno-Romero, J. A., Muñoz-Santos, C., Petrillo, G., Segura, S., Soyer, H. P., Zanchini, R., and Malvehy, J., 'Dermoscopy improves accuracy of primary care physicians to triage lesions suggestive of skin cancer,' Journal of Clinical Oncology, 2006, **24**(12), pp. 1877–82, ISSN 1527-7755, doi:10.1200/JCO.2005.05.0864.

[43] Siegel, R. L., Miller, K. D., and Jemal, A., 'Cancer statistics, 2016,' CA: A Cancer Journal for Clinicians, 2016, **66**(1), pp. 7–30, ISSN 00079235, doi:10.3322/caac.21332.

[44] Siegel, R. L., Miller, K. D., and Jemal, A., 'Cancer statistics, 2017,' CA: A Cancer Journal for Clinicians, 2017, **67**(1), pp. 7–30, ISSN 00079235, doi:10.3322/caac.21387.

[45] Friedman, R. J., Rigel, D. S., and Kopf, A. W., 'Early detection of malignant melanoma: the role of physician examination and self-examination of the skin.' CA: a cancer journal for clinicians, 1985, **35**(3), pp. 130–151, ISSN 0007-9235, doi:10.3322/canjclin.35.3.130.

[46] Rubegni, P., Feci, L., Nami, N., Burroni, M., Taddeucci, P., Miracco, C., Munezero Butorano, M. A. G., Fimiani, M., and Cevenini, G., 'Computer-assisted melanoma diagnosis: a new integrated system.' Melanoma research, 2015, **25**(6), pp. 537–42, ISSN 1473-5636, doi:10.1097/CMR.0000000000000209.

[47] Andreassi, L., Perotti, R., Rubegni, P., Burroni, M., Cevenini, G., Biagioli, M., Taddeucci, P., Dell'Eva, G., and Barbini, P., 'Digital dermoscopy analysis for the differentiation of atypical nevi and early melanoma: a new quantitative semiology,' Archives of Dermatology, 1999, **135**(12), pp. 1459–65, ISSN 0003-987X.

[48] Landau, M., Matz, H., Ethel, T., Dvir, M., and Brenner, S., 'Computerized system to enhance the clinical diagnosis of pigmented cutaneous malignancies,' International Journal of Dermatology, 1999, **38**(6), pp. 443–446, ISSN 00119059, doi:10.1046/j.1365-4362.1999.00629.x.

[49] Umbaugh, S. E., Moss, R. H., and Stoecker, W. V., 'Automatic Color Segmentation of Images with Application to Detection of Variegated Coloring in Skin Tumors,' IEEE Engineering in Medicine and Biology Magazine, 1989, **8**(4), pp. 43–50, ISSN 07395175, doi:10.1109/51.45955.

[50] Green, A., Martin, N., Pfitzner, J., O'Rourke, M., and Knight, N., 'Computer image analysis in the diagnosis of melanoma.' Journal of the American Academy of Dermatology, 1994, **31**(6), pp. 958–64, ISSN 0190-9622.

[51] Seidenari, S., Burroni, M., Dell'Eva, G., Pepe, P., and Belletti, B., 'Computerized evaluation of pigmented skin lesion images recorded by a videomicroscope: comparison between polarizing mode observation and oil/slide mode observation,' Skin Research and Technology, 1995, **1**(4), pp. 187–191, ISSN 0909-752X, doi:10.1111/j.1600-0846.1995.tb00042.x.

[52] Aitken, J. F., Pfitzner, J., Battistutta, D., O'Rourke, P. K., Green, A. C., and Martin, N. G., 'Reliability of computer image analysis of pigmented skin lesions of Australian adolescents,' Cancer, 1996, **78**(2), pp. 252–257, ISSN 0008-543X, doi:10.1002/(SICI)1097-0142(19960715)78:2<252::AID-CNCR10>3.0.CO;2-V.

[53] Ercal, F., Chawla, A., Stoecker, W. V., Lee, H. C., and Moss, R. H., 'Neural network diagnosis of malignant melanoma from color images,' IEEE Transactions on Biomedical Engineering, 1994, **41**(9), pp. 837–845, ISSN 00189294, doi:10.1109/10.312091.

[54] Ganster, H., Pinz, P., Rohrer, R., Wildling, E., Binder, M., and Kittler, H., 'Automated melanoma recognition,' IEEE Transactions on Medical Imaging, 2001, **20**(3), pp. 233–239, ISSN 02780062, doi:10.1109/42.918473.

[55] Heckbert, P., 'Color image quantization for frame buffer display,' in 'Proceedings of the 9th annual conference on Computer graphics and interactive techniques - SIGGRAPH '82,' volume 16, ACM Press, New York, New York, USA, ISBN 0897910761, ISSN 0097-8930, 1982 pp. 297–307, doi:10.1145/800064.801294.

[56] Faziloglu, Y., Stanley, R. J., Moss, R. H., Van Stoecker, W., and McLean, R. P., 'Colour histogram analysis for melanoma discrimination in clinical images,' Skin Research and Technology, 2003, **9**(2), pp. 147–56, ISSN 0909-752X, doi:10.1034/j. 1600-0846.2003.00030.x.

[57] Chen, J., Stanley, R. J., Moss, R. H., and Van Stoecker, W., 'Colour analysis of skin lesion regions for melanoma discrimination in clinical images,' Skin Research and Technology, 2003, **9**(2), pp. 94–104, ISSN 0909-752X, doi:10.1034/j.1600-0846. 2003.00024.x.

[58] Stanley, R. J., Stoecker, W. V., and Moss, R. H., 'A relative color approach to color discrimination for malignant melanoma detection in dermoscopy images,' Skin Research and Technology, 2007, **13**(1), pp. 62–72.

[59] Stanley, R. J., Moss, R. H., Van Stoecker, W., and Aggawal, C., 'A fuzzy-based histogram analysis technique for skin lesion discrimination in dermatology clinical images,' Computerized Medical Imaging and Graphics, 2003, **27**(5), pp. 387–396, ISSN 08956111, doi:10.1016/S0895-6111(03)00030-2.

[60] Barata, C., Ruela, M., Francisco, M., Mendonca, T., and Marques, J. S., 'Two systems for the detection of melanomas in dermoscopy images using texture and color features,' IEEE Systems Journal, 2014, **8**(3), pp. 965–979, ISSN 19379234, doi:10.1109/JSYST. 2013.2271540.

[61] Stoecker, W. V., *Computer applications in dermatology*, Igaku-Shoin New York, 1993.

[62] Nie, D., 'Classification of melanoma and clark nevus skin lesions based on Medical Image Processing Techniques,' in 'ICCRD2011 - 2011 3rd International Conference on Computer Research and Development,' volume 3, IEEE, ISBN 9781612848372, 2011 pp. 31–34, doi:10.1109/ICCRD.2011.5764239.

[63] Argenziano, G., Soyer, H. P., De Giorgi, V., Piccolo, D., Carli, P., and Delfino, M., 'Interactive atlas of dermoscopy (Book and CD-ROM),' 2000.

[64] McLean, R. P., *Tumor classification based on relative color analysis of melanoma and non-melanoma tumor images*, Ph.D. thesis, 1994.

[65] Stoecker, W. V., Li, W. W., and Moss, R. H., 'Automatic detection of asymmetry in skin tumors,' Computerized Medical Imaging and Graphics, 1992, **16**(3), pp. 191–197, ISSN 08956111, doi:10.1016/0895-6111(92)90073-I.

[66] Hance, G., Umbaugh, S., Moss, R., and Stoecker, W., 'Unsupervised color image segmentation: with application to skin tumor borders,' IEEE Engineering in Medicine and Biology Magazine, 1996, **15**(1), pp. 104–111, ISSN 07395175, doi:10.1109/51. 482850.

[67] KLIR, G. J. A. and FOLGER, T. A. A., *Fuzzy Sets, Uncertainty, and Information*, Prentice Hall, 1988, ISBN 9780133459845.

[68] Joe Stanley, R., Stoecker, W. V., Moss, R. H., Rabinovitz, H. S., Cognetta, A. B., Argenziano, G., and Peter Soyer, H., 'A basis function feature-based approach for skin lesion discrimination in dermatology dermoscopy images,' Skin Research and Technology, 2008, **14**(4), pp. 425–435, ISSN 0909752X, doi:10.1111/j.1600-0846.2008.00307.x.

[69] Dalal, A., Moss, R. H., Stanley, R. J., Stoecker, W. V., Gupta, K., Calcara, D. A., Xu, J., Shrestha, B., Drugge, R., Malters, J. M., and Perry, L. A., 'Concentric decile segmentation of white and hypopigmented areas in dermoscopy images of skin lesions allows discrimination of malignant melanoma,' Computerized Medical Imaging and Graphics, 2011, **35**(2), pp. 148–154, ISSN 08956111, doi:10.1016/j.compmedimag.2010.09.009.

[70] Rader, R. K., Payne, K. S., Guntupalli, U., Rabinovitz, H. S., Oliviero, M. C., Drugge, R. J., Malters, J. J., and Stoecker, W. V., 'The Pink Rim Sign : Location of Pink as an Indicator of Melanoma in Dermoscopic Images,' Journal of skin cancer, 2014, **2014**, p. 719740, ISSN 2090-2905, doi:10.1155/2014/719740.

[71] Braun, R. P., Gaide, O., Oliviero, M., Kopf, A. W., French, L. E., Saurat, J. H., and Rabinovitz, H. S., 'The significance of multiple blue-grey dots (granularity) for the dermoscopic diagnosis of melanoma,' British Journal of Dermatology, 2007, **157**(5), pp. 907–913, ISSN 00070963, doi:10.1111/j.1365-2133.2007.08145.x.

[72] Stoecker, W. V., Wronkiewiecz, M., Chowdhury, R., Stanley, R. J., Xu, J., Bangert, A., Shrestha, B., Calcara, D. A., Rabinovitz, H. S., Oliviero, M., Ahmed, F., Perry, L. A., and Drugge, R., 'Detection of granularity in dermoscopy images of malignant melanoma using color and texture features,' Computerized Medical Imaging and Graphics, 2011, **35**(2), pp. 144–147, ISSN 08956111, doi:10.1016/j.compmedimag.2010.09.005.

[73] Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D. M., Forman, D., and Bray, F., 'Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012,' International Journal of Cancer, 2015, **136**(5), pp. E359–E386, ISSN 00207136, doi:10.1002/ijc.29210.

[74] Egner, J. R., 'AJCC Cancer Staging Manual,' JAMA, 2010, **304**(15), p. 1726, ISSN 0098-7484, doi:10.1001/jama.2010.1525.

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

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

[77] Park, S. C., Tan, J., Wang, X., Lederman, D., Leader, J. K., Kim, S. H., and Zheng, B., 'Computer-aided detection of early interstitial lung diseases using low-dose CT images,' Physics in medicine and biology, 2011, **56**(4), pp. 1139–53, ISSN 1361-6560, doi:10.1088/0031-9155/56/4/016.

[78] MathWorks, 'Matlab Profiler,' 2012.

[79] Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P., *Molecular Biology of the Cell*, Garland Science, 4th edition, 2002.

[80] Guo, P., Almubarak, H., Banerjee, K., Stanley, R. J., Long, R., Antani, S., Thoma, G., Zuna, R., Frazier, S. R., Moss, R. H., and Stoecker, W. V., 'Enhancements in localized classification for uterine cervical cancer digital histology image assessment,' Journal of Pathology Informatics, 2016, **7**(1), p. 51, ISSN 2153-3539, doi:10.4103/2153-3539. 197193.

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

[82] Almubarak, H., Stanley, R. J., Long, R., Antani, S., Thoma, G., Zuna, R., and Frazier, S., 'Convolutional Neural Network Based Localized Classification of Uterine Cervical Cancer Digital Histology Images,' in 'Complex Adaptive Systems 2017 (Complex Adaptive Systems 2017),' volume 114, ISSN 18770509, 2017 pp. 281–287, doi:10.1016/j.procs.2017.09.044.

VITA

Haidar A. Almubarak was born in the city of Alahsa in the Eastern Provence, Saudi Arabia. He did his schooling at Almarkaz schools. He received his Bachelor degree in computer engineering from King Fahd University of Petroleum and Minerals in 2005. He worked at Saad Specialist Hospital in Alkhobar, Saudi Arabia until March of 2008, and then moved to work at Saudi Arabian Basic Industries Corporation (SABIC) in Jubail, Saudi Arabia. In September of 2009, he moved to the United States to pursue his higher education degrees. In 2011, he received his master degree in computer, information, and network security from DePaul University. There he worked as a network lab assistant helping instructors in preparing laboratory work for graduate and undergraduate students and helping students in their lab work. In 2014, he moved to Rolla, Missouri to pursue a Ph.D. in computer engineering where he worked as a research assistant and as a teaching assistant. He received a Ph.D. in computer engineering from Missouri University of Science and Technology in May 2018.