Spring 2019

Classification of Hyperspectral Colon Cancer Images Using Convolutional Neural Networks

Sean Joseph Mobilia
San Jose State University

CLASSIFICATION OF HYPERSPECTRAL COLON CANCER IMAGES USING CONVOLUTIONAL NEURAL NETWORKS

A Thesis

Presented to

The Faculty of the Department of Electrical Engineering

San José State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

by

Sean J. Mobilia

May 2019

The Designated Thesis Committee Approves the Thesis Titled

CLASSIFICATION OF HYPERSPECTRAL COLON CANCER IMAGES USING CONVOLUTIONAL NEURAL NETWORKS

by

Sean J. Mobilia

APPROVED FOR THE DEPARTMENT OF ELECTRICAL ENGINEERING

SAN JOSÉ STATE UNIVERSITY

May 2019

Birsen Sirkeci, Ph.D.	Department of Electrical Engineering
Robert Morelos-Zaragoza, Ph.D.	Department of Electrical Engineering
Chang Choo, Ph.D.	Department of Electrical Engineering

ABSTRACT

CLASSIFICATION OF HYPERSPECTRAL COLON CANCER IMAGES USING CONVOLUTIONAL NEURAL NETWORKS

by Sean J. Mobilia

Hyperspectral images are 3-D images, which contain data in hundreds of spectral bands as opposed to 2-D images, which contain data in at most 3 bands (red, green, and blue). Hyperspectral imagery was initially developed for remote sensing; however, recently, researchers have started to see its potential in medical diagnosis and cancer detection. Hyperspectral images provide massive amounts of data about the objects they are studying, and this causes challenges during information processing. Machine learning tools, such as convolutional neural networks (CNNs), are known to be successful in extracting features and classifying traditional 2-D images. This thesis proposes CNN architectures for processing hyperspectral data for colon cancer detection. Using data taken from a limited number of colon tissue samples, this thesis shows that the proposed CNN architecture can classify cancerous and noncancerous tissue samples utilizing hyperspectral information. The obtained results are compared to grayscale images of the same tissue samples, looking both at grayscales of the individual hyperspectral bands and panchromatic grayscale images in which the spectral bands are merged together. The CNN using the hyperspectral data shows advantages over the grayscale data, with a 5.6% improvement in accuracy and a 0.037 improvement in F1 score over the individual band grayscale images and a 21.7% improvement in accuracy and a 0.178 improvement in F1 score over the panchromatic grayscale images. The results are also compared to a K-nearest neighbor (KNN) classifier and a logistic regression (LR) classifier using the hyperspectral data, and the CNN shows advantages over both. The CNN has a 17.9% improvement in accuracy and a 0.141 improvement in F1 score over the KNN classifier and a 5% improvement in accuracy and a 0.061 improvement in F1 score over the LR classifier.

ACKNOWLEDGMENTS

I would like to thank my advisor, Professor Birsen Sirkeci, for all her help and support during this project and for having tremendous patience with me. I would also like to thank my other committee members, Professor Robert Morelos-Zaragoza and Professor Chang Choo, for taking the time to read and review my work.

I would like to thank Dr. Paul Rider (Dept. of Surgery at University of South Alabama), Dr. Carole Boudreaux (Dept. of Pathology at University of South Alabama), and Dr. Shante Hill (Dept. of Pathology, University of South Alabama) for their assistance in obtaining the clinical specimens for this study. I would like to thank Dr. Silas J. Leavesley (University of South Alabama) and Dr. Thomas C. Rich (University of South Alabama) for allowing me to build upon their work and use their hyperspectral data. I would like to thank Joshua Deal (University of South Alabama), for taking the time to answer my questions and for working with me on using and understanding the hyperspectral dataset.

Last, but not least, I would like to thank my friends and family for putting up with me while I've been working on this thesis. I'll try to hang out more soon.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANN	Artificial Neural Network
CNN	Convolutional Neural Network
DNN	Deep Neural Network
ELU	Exponential Linear Units
FC	Fully Connected Layer
Hyper	Hypercube Dataset
IB	Individual Band Dataset
KNN	K-Nearest Neighbors
LOPO	Leave One Patient Out
LR	Logistic Regression
PC	Panchromatic Dataset
PCA	Principal Component Analysis
RELU	Rectified Linear Units
SPC	Segmented Panchromatic Dataset
SH	Segmented Hypercube Dataset
SSPC	Smaller Segmented Panchromatic Dataset
SSH	Smaller Segmented Hypercube Dataset
UAV	Unmanned Aerial Vehicle

1 INTRODUCTION

Across the globe, cancer is one of the leading causes of death among all age groups. It can affect anyone, and just about every person has either had it, or known someone who has had it. Among people in the United States, colon cancer is the second largest cause of cancer deaths [1]. Colon cancer symptoms can be difficult to detect at first, as many of the early symptoms may seem like common ailments. Colon cancer is diagnosed in the localized stage, the earliest stage in which the cancer is smallest, 38% of the time. It is diagnosed in the regional stage, in which the cancer has spread throughout the organ, 36% of the time. And it is diagnosed in the distant stage, in which the cancer has spread throughout the body into other organs, 22% of the time [2]. Fig. 1 shows the stages of colon cancer when it is first diagnosed.

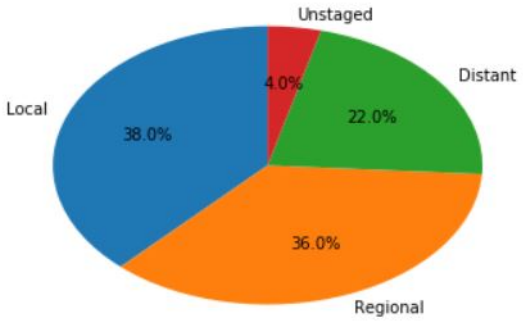

Fig. 1. Colon cancer stage at diagnosis. Data are provided by Siegel et al. [2].

As with many forms of cancer, early screening and diagnosis significantly increase the patient's chance of survival. Fig. 2 shows colon cancer survival rates at different stages of diagnosis. If caught early, the 5-year survival rate for colon cancer can be as high as 91.1%; however as the cancer spreads, the rate drops significantly. If cancer is diagnosed in the distant stage, the survival rate falls to 13.3% [2]. Current screening methods are missing too many cases until the cancer has spread into other parts of the body, so an

improvement to these methods is needed. One of the areas being explored is the use of hyperspectral imaging to aid in early cancer detection and research [3].

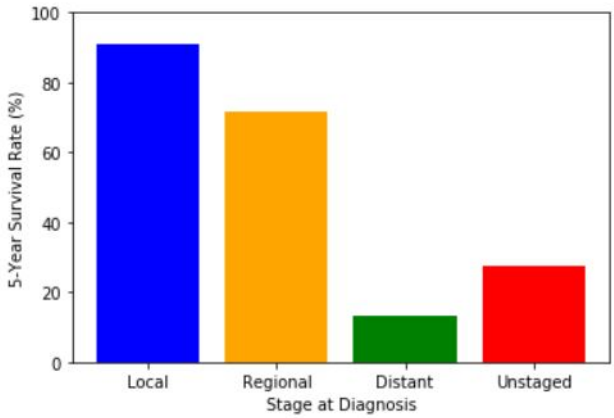

Fig. 2. Colon cancer survival rates by stage detected. Data are provided by Siegel et al. [2].

Hyperspectral imaging involves imaging data that take a detailed look at the electromagnetic spectrum in order to learn more detail about the subject being studied. Initially developed for remote sensing satellite instruments, hyperspectral imaging provides a wealth of information about the material composition of the subject being studied [4]. With advancements in sensor size and computation power, this imaging is finding more applications in laboratory research and commercial industries.

Hyperspectral images contain a tremendous amount of information. Many typical hyperspectral imagers create hundreds of spectral bands with each image. This huge amount of data can be difficult to process and examine with traditional image processing tools. Artificial neural networks (ANNs) can be used to extract the relevant information from this complex dataset. First developed in the 1940s, ANNs are programs modeled after brains [5]. ANNs involve the use of hidden computational layers to extract deeper features and relationships between different inputs. These hidden layers allow ANNs to move beyond basic tasks [6]. ANNs that have two or more of these hidden layers are

known as deep neural networks (DNNs). The popularity of these networks has been growing recently due to the DNN's ability to handle increasingly complex tasks, such as speech recognition, prediction of molecule interaction, image segmentation, and classification, thanks to the advances in computational tools [6], [7]. Convolutional neural networks (CNNs) are special kinds of DNNs that make use of discrete convolutions to process large datasets and extract features from them. These networks are especially useful when it comes to image classification, due to the convolutional layers picking out abstract features independently of where they are located in the image.

The purpose of this thesis is to demonstrate the use of CNNs in processing hyperspectral imaging data for classifying colon cancer images. This will be accomplished by training a CNN to classify hyperspectral colon tissue samples and then comparing the results to a similar CNN trained on grayscale images generated by the same test samples. The output from the CNN will also be compared to the output from simple classifiers, to show that deep learning tools were beneficial in classifying the data. The remainder of this thesis is structured as follows. Chapter 2 will cover some necessary background related to hyperspectral imaging and CNNs. Chapter 3 will introduce some of the latest trends in cancer detection, including some applications making use of hyperspectral technology. Chapter 4 will discuss how the data used in this study were collected. Chapter 5 will present how the CNN was developed and implemented. Chapter 6 will present the results of the study and discuss the performance of the CNN. In Chapter 7, we will discuss the future work. Finally, in Chapter 8, we will present the conclusions.

2 BACKGROUND

This chapter presents the necessary background for hyperspectral imaging, beginning with an overview and explanation of what hyperspectral imaging is. The chapter will then discuss some of the methods for processing hyperspectral imaging data that are available, and then it will discuss some of the applications of hyperspectral imaging data and why they are relevant. This chapter also presents the relevant background on CNNs, discussing their development, key features, and implementation.

2.1 What is a Hyperspectral Image?

Humans are able to visualize only a small part of the electromagnetic spectrum. The light emitted by different objects and materials has its own electromagnetic signature, and since the 1960s, sensors have been used that simultaneously gather information on both the visible and invisible parts of the spectrum to better collect these signatures. Hyperspectral imaging involves sensors that look at several narrow bands of the electromagnetic spectrum [8]. This allows the sensor to catch subtle shifts that are not noticed by traditional imaging systems. These bands can range between 2 – 10 nm each and can cover the ultraviolet, visual, near-infrared and short-wave infrared portions of the spectrum. These images tend to be both spectrally and spatially smooth (the neighboring pixels and the neighboring bands tend to be highly correlated) [4]. A visual representation of these bands, called a hypercube, is shown in Fig. 3 using the colon cancer imaging data discussed in Chapter 4. The light captured by a hyperspectral imager is usually made up of a collection of constituent spectra, called endmembers. These spectra get mixed together by a combination of atmospheric effects, spatial resolution of the imager, or the presence of gaseous mixtures in the medium [4]. The endmembers represent the pure materials captured in the image [9]. The materials considered a pure material in an image may vary based on the subject being studied. Highly reflective materials can also change

what proportions of endmembers are actually being measured. This can add to the challenge and complexity of analyzing hyperspectral data.

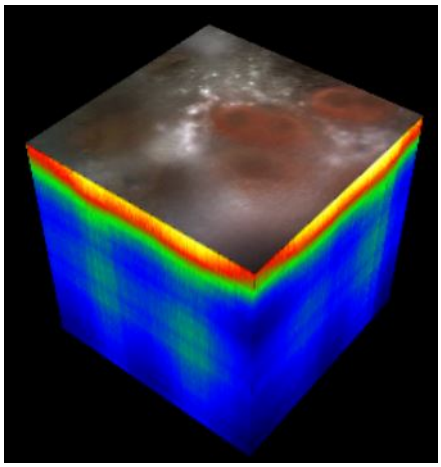

Fig. 3. Color representation of hyperspectral cube.

2.1.1 Processing of Hyperspectral Images

The high dimensionality and large amounts of data associated with hyperspectral images create a unique set of challenges for processing and analysis. This section will discuss some of the methods that have been developed to handle hyperspectral images, such as hyperspectral data fusion, spectral unmixing, and classification.

Data fusion is the processing technique in which spatial and spectral information is fused with one or more sources [4]. These techniques are generally implemented in order to improve the spatial resolution of an image (or at least clear up the noise). Data fusion techniques tends to be used more with hyperspectral satellite imagery, as there would be other available datasets to merge with.

There are three main types of hyperspectral data fusion: restoration, spectral data fusion, and spatial data fusion. Restoration involves the extension of traditional image de-noising techniques by taking advantage of the high spectral redundancy in the spectral bands. This merely preserves the original spatial and spectral resolutions and does not

improve upon them. Spectral data fusion involves fusing together hyperspectral bands, in order to reduce some of the inherent redundancy in the system. Spatial data fusion involves methods of improving the spatial resolution of an image by inferring information at the subpixel level [4]. This can be done by either subsampling the pixels in the image or by fusing the image with another image of higher spatial resolution.

Hyperspectral unmixing refers to process in which the pixel spectra are separated into endmember signatures [9]. These processes are trying to untangle different collected spectra into their pure components. Unmixing methods are often dependent on assumptions about the object or area being imaged. The linear unmixing model assumes that endmember abundances are associated with a percentage of the material the endmember represents. This tends to hold when the system being monitored is macroscopic. Linear mixing tends to occur in the hyperspectral instrument due to low spatial resolution. Nonlinear mixing tends to occur due to interactions in the atmosphere, occurring when light is reflected off of an object [9].

Hyperspectral classification is another common way to process hyperspectral imagery. One of the key aspects of hyperspectral classification is feature mining, the ability to extract features from the fine details. Some typical classification tools for hyperspectral imagery involve principal components analysis (PCA), support vector machines, linear discriminant analysis and quadratic discriminant analysis [4].

2.1.2 Applications of Hyperspectral Images

Advancements in technology have allowed hyperspectral imaging to expand beyond remote sensing satellites. Smaller camera sizes and more efficient storage have opened up the potential of unmanned aerial vehicle (UAV) and ground based commercial and research opportunities involving hyperspectral imaging. Precision agriculture is one of the potential applications hyperspectral imaging. With access to UAVs, farmers can use hyperspectral images to better monitor their crops. Hyperspectral images can help detect

disease, water stress levels, crop nutrients and invasive insects, as well as nutrient deficiencies and soil contaminants. This allows farmers to catch these problems well before the first obvious visual symptoms appear [8]. Fresh water resource management is another potential application of hyperspectral imaging. Hyperspectral imaging can help provide accurate estimates of the available ground water and can help provide accurate early warning and flood monitoring. It can also provide information on the quality of water available. Hyperspectral imaging can be used to catch any biological or biochemical contaminant and can measure the chemical oxygen and concentration of chlorophyll in the water [8].

Hyperspectral imaging also has many defense and homeland security related applications. Hyperspectral anomaly detection methods can easily distinguish between targets and the background. On the ground, hyperspectral imaging can be used as a tool for facial recognition, using a combination of classification techniques such as support vector machines and linear unmixing [8]. It also can be used in facial stress detection, by detecting changing blood oxygen levels in a subject.

Hyperspectral imaging is also finding more use on the ground, in applications involving food safety and quality. The rich amount of spatial and spectral information present in these images make them ideal for analyzing food quality and detecting spoilage and contaminants. Hyperspectral imaging has been used to assess the firmness of blueberries and the fat content of pork chops [8]. Hyperspectral imaging also has art and forensic applications. The spectral signatures can be helpful in detecting forgeries and can help to authenticate the value of certain pieces [8]. Hyperspectral imaging also has many medical diagnostic based applications, such as bruise and burn detection, cancer screening, and surgery assistance [10]. Hyperspectral imaging has also been shown to be useful in detecting bleeding, diabetes, and several forms of cancer. It is being tested as a potential alternative to computed tomography and magnetic resonance imaging [8].

2.2 A Brief Overview of CNNs

The development of CNNs was inspired based on studies of the visual cortex of mammals [7]. These studies showed that each of the visual cortex neurons reacted to only a limited portion of the animal's field of vision. These neurons would react differently to different visual features. These studies paved the way for the architecture of LeNet-5, developed in LeCun et al. [11]. This architecture is implemented by using a combination of convolution layers, which extract features from the input, and pooling layers, which subsample the input. A visual representation of a CNN architecture is shown in Fig. 4. From left to right in the figure, the input is shown having the features extracted by the various layers of the network.

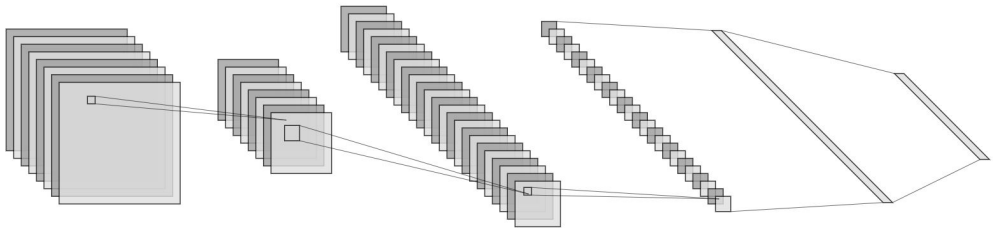

Fig. 4. Visualization of CNN architecture for proposed hyperspectral classifier.

CNN architecture makes use of the properties of sparse interactions, local receptive fields, parameter sharing, equivarient representation, and sub-sampling to create these powerful models [6], [11]. The property of sparse interactions means that the weights of each convolution layer are not fully connected to the layers being fed into them. CNNs employ sparse interactions by making use of kernels with smaller sizes than the inputs. This leads to fewer weights and allows the networks to be programmed more efficiently. Sparse interactions tie into the property of local receptive fields, in which each output receives inputs from a small portion of neighboring inputs. This allows the subsequent layers to extract deeper and more abstract features from the input. Another useful property of CNNs is parameter sharing. Typically, in ANNs, each weight is used by one element of

the input. For large images, this would lead to rather costly networks. However, for CNNs, the weights related to one portion of the input are reused for each of the other inputs. Combined with sparse interactions, these two properties help reduce the size of CNNs and make them into a valuable tool for learning larger inputs. The property of equivarient representation, in which changes to the input lead to similar changes in the output, is also an important property of CNNs. This property is useful for extracting certain features, such as edges in an image [6]. Sub-sampling helps to further reduce the size of the feature maps and makes CNNs much more efficient [11].

As CNNs are based on the visual cortex, they naturally excel at visual tasks. As well as being used for self-driving cars, image classification, and other visual tasks, CNNs have been also used in voice recognition and natural language processing applications [7].

CNNs are made up of a number of convolutional layers, which are sparsely connected to the previous layer by a kernel. This limited connection over a local field is what allows CNNs to extract features from the input data [7]. This kernel sweeps over the input to the convolution layer, and in doing so, the convolution layer implements 2-D convolution on the input layer. This is given as

$$S(i,j) = (K * I)(i,j) = \sum_m \sum_n I(i-m, j-n)K(m,n). \tag{1}$$

Many machine learning algorithms actually implement a related function, the cross-correlation function, to avoid having to flip the kernel. The cross-correlation function is given as

$$S(i,j) = (K * I)(i,j) = \sum_m \sum_n I(i+m, j+n)K(m,n). \tag{2}$$

In practice this is acceptable because the machine learning algorithm will just learn the values of the flipped kernel [6]. The Python module Tensorflow uses cross-correlation to implement convolution in its functions.

The convolution layer is generally followed by an activation function. Some common activation functions are shown in Fig. 3. Typically the relu function is the most popularly used activation function; however some networks such as LeNet-5 have been known to use the tanh function [11].

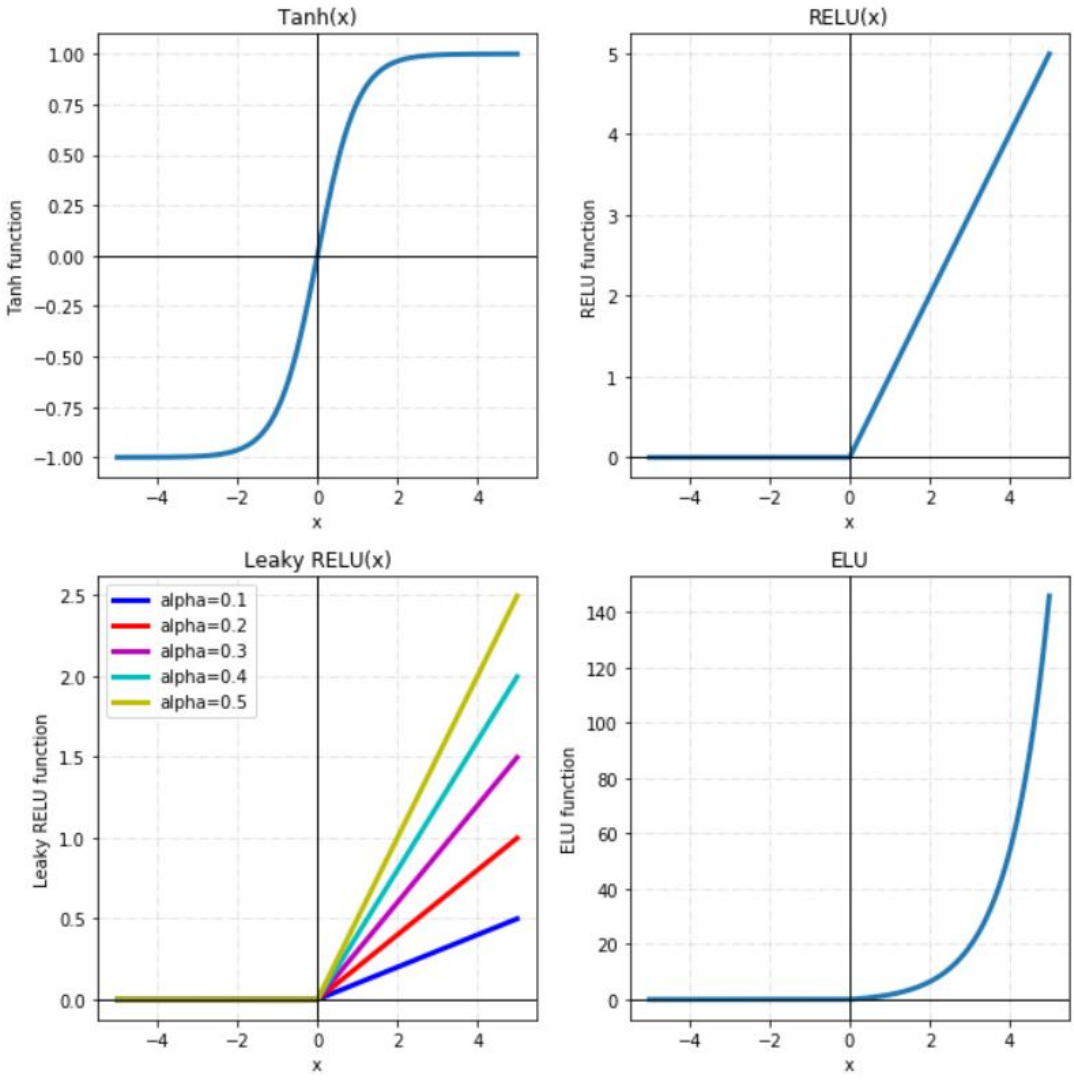

Fig. 5. Common activation functions used in CNNs.

The other main part of a CNN is the pooling layer. The goal of these layers is to shrink the input down to prevent overfitting on the network. The kernel on these layers

has no weights and uses an aggregation function to shrink the size of the input. Most CNNs make use of max pool layers, which take the maximum value at the kernel as it passes over the input. The remaining inputs are dropped [7].

3 LITERATURE REVIEW

This chapter will provide a background of previous work done in hyperspectral imagery and machine learning tools related to cancer detection. First, the chapter will discuss machine learning tools and their use in cancer detection. The chapter will then discuss hyperspectral imagery and machine learning tools used in tandem to detect cancer cells.

3.1 Machine Learning and Cancer Detection

The most popular machine learning tools for cancer detection are support vector machines, random forests, Bayesian networks, and CNNs. Bazazeh and Shubair [12] showed that mammogram images can be used with support vector machines and random forests to detect breast cancer. Alam et al. [13] used computerized tomography scans with support vector machines to classify lung cancer.

CNNs have recently been applied as a tool in cancer diagnosis. Tan et al. [14] discussed using CNNs to help detect breast cancer using grayscale mammogram images. Godkhindi and Gowda [15] developed colon cancer detection methods with K-nearest neighbor (KNN) algorithms, random forests, and CNNs for use with grayscale computed tomography colonography images. Chiang et al. [16] used 3-D ultrasound images to improve breast cancer detection using a 3-D CNN.

3.2 Hyperspectral Imaging and Cancer Detection

As hyperspectral imaging is becoming more widespread in commercial and research applications, it is finding new uses as a tool for medical diagnosis. Lu and Fei [10] have discussed the use of hyperspectral imagery and its role in medical diagnostics and as a tool for image-guided surgery. They also discussed several hyperspectral systems that have been develop to detect and screen for different types of cancer, such as skin, breast, gastric, cervical, and prostate cancers. Many of these cancer detection applications focus

on the visible parts of the spectrum, especially focusing on the blue and green wavelengths. It has been shown in Leavesley et al. [17] that for colon cancer, the blue-violet portion of the spectrum (390 nm – 450 nm) has the highest correlation with instances of cancer.

Machine learning and hyperspectral images have been combined to create a system of new, noninvasive forms of cancer detection [3], [10], [17]–[23]. Support vector machines and minimum spanning forests have been used to analyze hyperspectral data [19]–[21] with PCA having been used for feature reduction to improve the algorithms [19]. Rathore et al. [3] discussed methods for hyperspectral colon cancer detection using support vector machines, PCAs, and linear discriminant analysis. Lu et al. [18] used a combination of spatial and spectral data to improve non-invasive head and neck cancer detection techniques, using support vector machines and PCAs. Pike et al. [20] used tumor-bearing mice to test out the hyperspectral classification of head and neck cancer using minimum spanning forests. Gopi and Reshmi [21] developed a method combining minimum spanning forests with support vector machines, and applied it in order to detect cancer on various animals.

Hyperspectral imaging has also been increasingly used in conjunction with neural network models to improve forms of cancer detection. Along with using support vector machines, Leavesley et al. [22] has used an ANN to classify colon cancer images using hyperspectral fluorescence. Nathan et al. [19] has also explored using an ANN with hyperspectral imagery to classify cancer data. Ma et al. [23] have used CNNs in the pixel classification of hyperspectral images in order to detect head and neck cancer. They turned the spectra of the image from each pixel into a 16×16 patch and fed this into a CNN.

4 HYPERSPECTRAL COLON CANCER DATA

In this chapter we discuss the hyperspectral data that are used in this research. In the first section, we discuss how the colon tissue samples were gathered. In the next section, we discuss the imager that was used to gather the data. In the final section of the chapter, we discuss the distribution of the data that were collected.

The colon cancer tissues samples that are used in this research were gathered as previously described in Deal et al. [24] from the University of South Alabama Medical Center Department of Surgery. This dataset is not publicly available and is used as part of a collaboration with researchers at the University of South Alabama. During the surgery, the tumorous growth was removed along with a section of normal tissue. The samples were then taken to pathology where they were bisected, with one half frozen and analyzed and the other half stored for imaging. The half stored for imaging were returned in cups marked “lesional” and “normal” and then rinsed and cut into 2 cm cubes. Each of the samples was imaged within 8 hours of being resectioned. An example of one of these tissue samples is shown in Fig. 6. Each tissue sample had images taken from at least two fields of view at the lab.

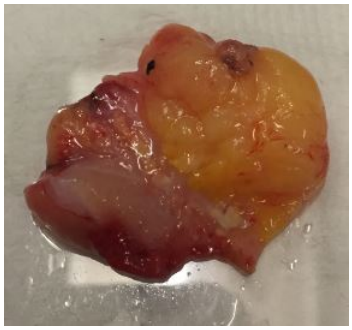

Fig. 6. Example of tissue sample studied with the hyperspectral imager. Image provided by University of South Alabama researchers.

The collected colon tissue samples were imaged with QImaging Corporation’s Rolera EM-C^2 camera at the laboratory, using the excitation wavelength tuning method described

in Deal et al. [24]. The camera has a 14-bit digital output, outputting a unitless spectral intensity at each pixel with a range between 0 and 16383 (2^{14}-1). Each pixel has a size of 8 μm $\times$ 8 μm. The hyperspectral filters were tuned to the wavelengths from 360 nm to 550 nm, with a spectral resolution of 5 nm. The band corresponding to the 385 nm window was not present for each dataset, and has been excluded in this research. The total number of spectral bands in each hypercube is 38.

The University of Southern Alabama was able to provide 175 hyperspectral tissue sample images, taken from 13 patients, for this thesis. These images all have a size of 501 $\times$ 502 $\times$ 38 (for a total of 9,557,076 pixels or features per image). There are 88 images labeled as lesional tissue samples and there are 87 images labeled as nonlesional tissue samples. The breakdown of these sample hyperspectral tissue images is shown in Table 1.

Table 1

Breakdown of Tissue Sample Images by Patient

Patient ID	Number of Lesional Images	Number of Nonlesional Images	Total
1	3	2	5
2	5	4	9
3	3	5	8
4	8	6	14
5	9	9	18
6	4	6	10
7	15	6	21
8	9	6	15
9	7	8	15
10	8	12	20
12	10	9	19
13	5	6	11
14	2	8	10
Total	88	87	175

As can be seen from the table, there is a lot of variation in the number of images provided for each patient, with some patients having as few as 5 images, and others as many as 21.

And while the proportion of cancerous to noncancerous images varies from patient to patient, the total data are split almost evenly between the cancerous and noncancerous images.

There appears to be considerable spectral variation between patients as well. Fig. 7 shows the spectral distribution from lesional samples, with spectra of the same color coming from the same patient.

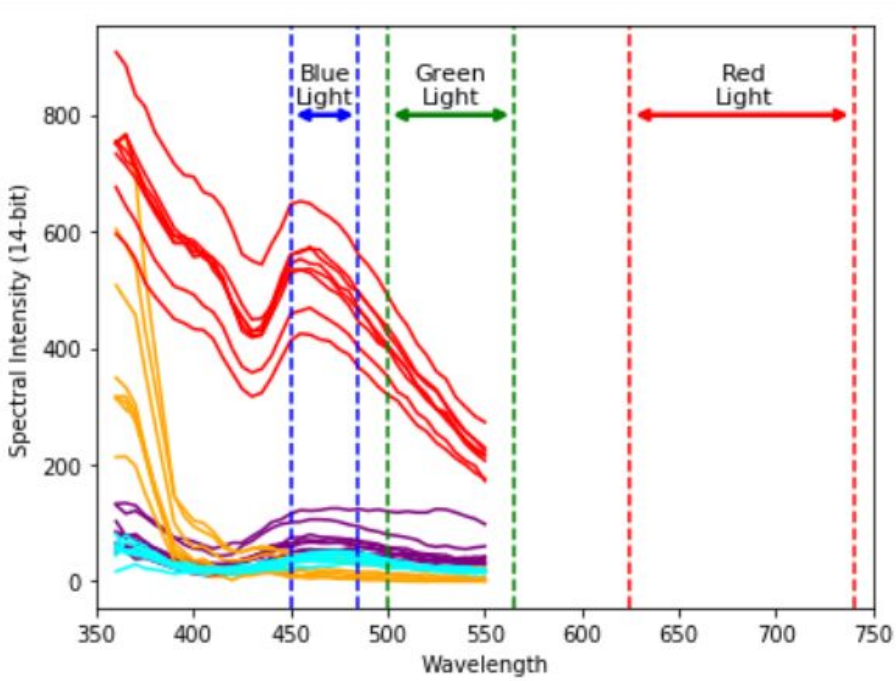

Fig. 7. Spectral distribution of lesional images, taken at random pixels.

Fig. 8 shows the spectral distribution of some nonlesional samples, again with spectra of the same color coming from the same patients. The dotted lines in these figures correspond to the wavelengths associated with blue, green, and red light. From these images, it can be seen that the spectral values of tissue samples tend to be highly correlated between patients, but also differ from patient to patient.

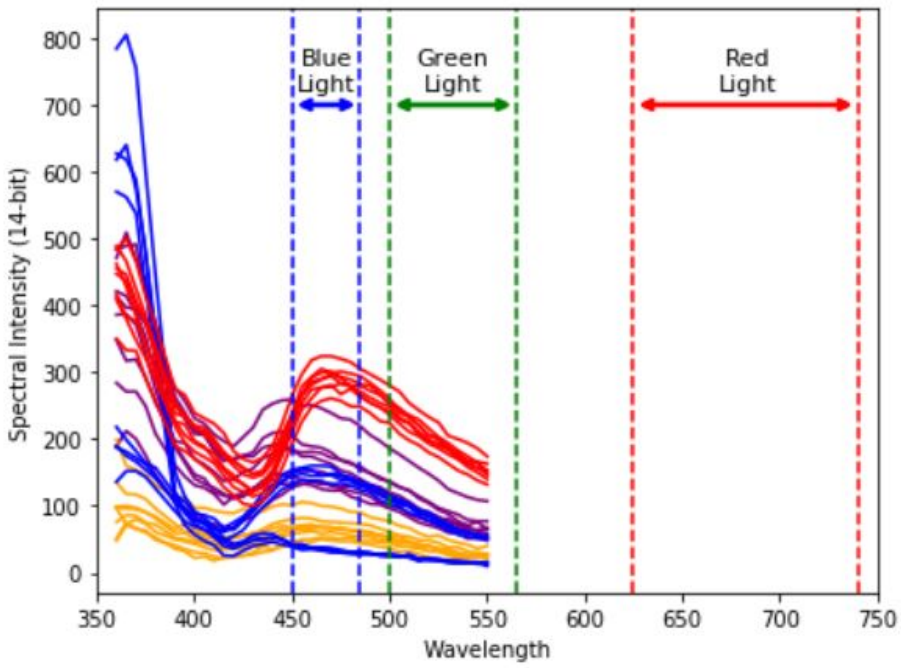

Fig. 8. Spectral distribution of nonlesional images, taken at random pixels.

5 PROPOSED CONVOLUTIONAL NEURAL NETWORK

This chapter discusses the development and implementation of the hyperspectral classifier. First the datasets being compared are discussed. The preprocessing of the data is then presented. The system architecture is presented, and finally the chapter discusses the classifier implementation.

5.1 Data Augmentation and Data Deduction

Three schemes have been developed with the hyperspectral data in order to test the advantages of the hyperspectral information in classifying these images. The first of these schemes involves merging all 38 of the 5 nm bands into one grayscale image. These are the panchromatic (PC) images, and this results in 175 single-channel images. The second scheme involves each of hyperspectral bands being treated as its own individual grayscale image. These are the individual band (IB) images, and this results in 6650 single-channel images. The third scheme involves the complete hypercube (Hyper) images. This group has 175 hyperspectral images, each with 38 channels. These schemes result in a large difference in the number of images between them. The PC and Hyper schemes use only the 175 images, while the IB scheme has 6650 images to train with.

One key assumption is made regarding the tissue samples and these schemes. The tissue samples were sent to the lab in dishes that were only labeled either “lesional” or “nonlesional.” From this it is assumed that each tissue sample is homogeneous. It is either completely ”lesional” or completely ”nonlesional.” Based on this assumption, the PC and Hyper schemes are each divided into two smaller sets. The first sets are the segmented hypercubes (SH) and the segmented panchromatic (SPC) schemes, each with 100×100 pixel images (4375 images total). The second of these sets are the smaller segmented hypercubes (SSH) and the smaller segmented panchromatic (SSPC) schemes, with 50×50 pixel images (17500 images total). These five schemes are summarized in Table 2, and these are used to test the proposed architecture.

Table 2
Feature Summary of Tissue Sample Schemes

Scheme	Total Number of Images	Image Dimensions	Number of Features
IB	6650	$501 \times 502 \times 1$	251502
PC	175	$501 \times 502 \times 1$	251502
Hyper	175	$501 \times 502 \times 38$	9557076
SPC	4375	$100 \times 100 \times 1$	10000
SH	4375	$100 \times 100 \times 38$	380000
SSPC	17500	$50 \times 50 \times 1$	2500
SSH	17500	$50 \times 50 \times 38$	95000

5.2 Data Preprocessing

The preprocessing of the image involves shuffling the dataset, normalizing the dataset, and flattening and storing the data in an hdf5 file. The patient order is shuffled by a random seed, and the images are grouped together by patient so the images training the network are from different patients than the ones validating the network. The patients are shuffled as shown in Table 3.

Table 3

Patient Order

Validation Fold	Patient ID
1	7
2	4
3	12
4	8
5	5
6	14
7	2
8	3
9	6
10	9
11	1
12	10
13	13

The dataset is then normalized by dividing all of the pixels by the maximum possible value of the dataset (determined by the maximum possible output of the 14-bit camera specifications, which is 16383 or $2^{14}-1$). These images are then flattened and stored in an hdf5 file before being loaded and trained to the CNN. The lesional images are labeled 0 and the nonlesional images are labeled 1.

5.3 CNN Architecture

These schemes are tested on a CNN made up of three alternating sets of 2-dimensional convolution layers and max pooling layers. These sets are followed by a fully connected layer which is connected to the output. This network uses L2 regularization and the convolutional layers carry out 2-dimensional convolutions (the convolution is only carried out along the x-axis and y-axis of the image). Each kernel has the same depth as the input to the convolution layer. The outputs of the convolution and fully connected layer all use tanh as the activation function. The basic CNN architecture is shown in Table 4.

Table 4

Basic CNN Architecture

CNN Layer	Kernel Window Size	Stride	Feature Maps
Conv+tanh+Drop	5 X 5 X d	3	24
Max Pool	2 X 2 X d	2	24
Conv+tanh+Drop	3 X 3 X d	1	48
Max Pool	2 X 2 X d	2	48
Conv+tanh+Drop	2 X 2 X d	1	64
Max Pool	2 X 2 X d	2	64
FC Layer	24	-	-

The first convolutional layer in the table is connected to the input and the fully connected layer is connected to the output of the network. The table also shows the two-dimensional filter size and the stride of the filter. Each of the convolutional layers uses zero padding on the inputs while the max pool layers uses no zero padding.

5.4 CNN Implementation

This CNN is implemented in Python using Tensorflow with an NVIDIA Tesla K40C GPU. Since the images from each patient are all from the same tissue sample, there is high correlation between the images from each patient and each segment. Before being fed to the network, the images are separated into 13 training and validation folds, and a variation of the leave one out model evaluation method is used. Leave one patient out (LOPO) uses the images from 12 of the 13 patients to train a CNN and then tests the model on the remaining patient's data.

These models are trained for a set number of epochs, which is determined by an initial test run for each scheme. The models are trained for enough epochs to ensure that the network has converged to a stable point. An example of the SSH scheme converging is shown in Fig. 9.

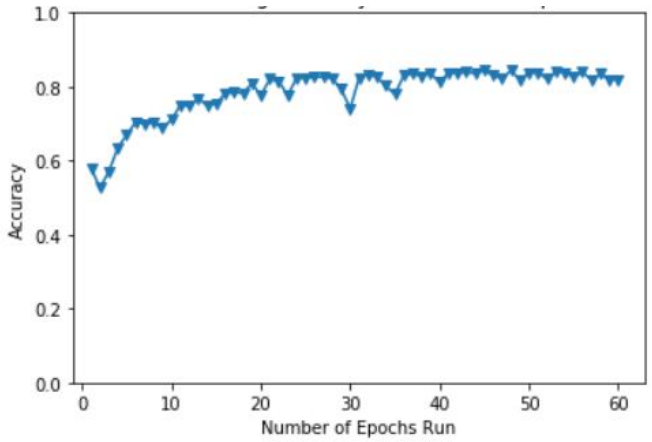

Fig. 9. Training accuracy per epoch trained for the SSH scheme.

The number of epochs each scheme is trained for is shown in Table 5. These models are evaluated on the basis of their accuracy, precision, recall, and F1 score. Accuracy measures the number of correct predictions over the total predictions made. Precision measures the number of true positives over the total number of positive predictions made.

Table 5
Training Epochs for Each Scheme

Scheme	Training Epochs Run
IB	30
PC	100
Hyper	80
SPC	120
SH	90
SSPC	120
SSH	60

For this dataset, a high precision means that fewer of the cancerous samples are predicted as noncancerous. Recall measures the number of true positives over the total number of positives. For this dataset, a high recall means that fewer noncancerous images are being classified as cancerous. The F1 score is the combination of the precision and recall, using the harmonic mean.

The performances of the last 20 epochs are averaged together and this is the metric used to evaluate each model. For the segmented schemes, the validation set segments are also grouped by the original image, and a final image prediction is made based on the majority of the segments from that image. A similar method is used for the IB scheme, in which each image has its bands grouped together and a final prediction is made based on the majority of the bands.

6 HYPERSPECTRAL CLASSIFIER PERFORMANCE

This chapter covers the performance of the CNN trained to classify the hyperspectral colon cancer images. The first section compares the different spectral schemes that have been used to determine the contribution of the hyperspectral data in classifying the lesional tissue samples. The next section compares the results from the CNN to different simple classifiers in order to determine the usefulness of the CNN in classifying the hyperspectral data. Then the chapter looks at the efforts to fine tune the hyperparameters of the CNN. The final section further discusses the classifier results.

6.1 Spectral Data Comparison Results

The goal of this thesis is to determine whether the hyperspectral data have been useful in classifying the colon cancer tissue samples. For this reason, the PC, IB, and Hyper images and their segmented counterparts (see Section 5.1 for details) are trained on the CNN architecture discussed in Section 5.3. The results of these models are shown in Table 6.

Table 6

LOPO Validation Set Results

Models	Acceleration	Precision	Recall	F1 Score
IB	0.603	0.609	0.575	0.590
PC	0.530	0.529	0.515	0.521
SPC	0.485	0.480	0.473	0.476
SSPC	0.442	0.439	0.460	0.449
Hyper	0.476	0.467	0.403	0.431
SH	0.630	0.640	0.588	0.612
SSH	0.659	0.689	0.577	0.627

Training using full-sized hyperspectral images does not perform better than the other sets of networks. However, in this case, the PC and Hyper schemes have a much smaller number of available images than the IB scheme. Once the images are segmented, then the hyperspectral data are an improvement when they are compared to the other data schemes.

The SSH scheme outperforms each of the other datasets with 65.9% accuracy and an F1 score of 0.627. We believe the overall poor performance of the classifiers is due to the small number of patients available for this study. Despite this limitation, the SSH scheme still outperforms the IB scheme by 5.6% (and by a 0.037 F1 score). This shows that the spectral information present in the hypercube helps determine whether or not the tissue is cancerous. Table 7 takes a look at the breakdown of the SSH classifier performance by patient.

Table 7

SSH Performance by Patient

Patient	Acceleration	Precision	Recall	F1 Score
1	0.510	0.092	0.125	0.105
2	0.811	0.936	0.588	0.708
3	0.375	0.000	0.000	0.000
4	0.814	0.990	0.575	0.721
5	0.650	0.890	0.339	0.473
6	0.880	1.000	0.800	0.887
7	0.295	0.209	0.642	0.315
8	0.600	0.000	0.000	0.000
9	0.870	0.880	0.875	0.873
10	0.778	0.990	0.638	0.772
12	0.584	0.655	0.639	0.606
13	0.672	0.638	0.958	0.763
14	0.830	1.000	0.788	0.879
Overall	0.659	0.689	0.577	0.627
μ	0.667	0.637	0.536	0.546
σ	0.181	0.394	0.309	0.321

As shown in the table, there is a lot of variation between patients. Some of this is caused by the number of images the tissue samples from each patient have provided. Another factor that affects the patient performance is the small number of patients available in the study. For two patients (patient 3 and patient 8), the model predicts that each image is lesional.

The model performs best on patient 6's data, with 88% accuracy and 0.887 F1 score. The model accuracies for the data for patients 2, 4, 9, and 14 are over 80%. On the other hand, the model accuracies for the data for patients 3 and 7 are below 50%. The results of these models vary highly from patient to patient. This is due to the small number of patients available for the study and the high variation between the spectra from each patient.

The confusion matrix from the SSH LOPO validation set is shown in Table 8. This classifier tends to misclassify slightly more nonlesional tissue samples than lesional ones. This observation is similar to the results shown in Leavesley et al. [22]. This might make the network more useful as a preliminary cancer screening, as the model tends to misclassify fewer cancerous samples.

Table 8
Confusion Matrix for the SSH LOPO Validation Set

SSH		
	Predicted Labels	
True Labels	Lesional	Nonlesional
Lesional	65	23
Nonlesional	37	50

6.2 Classifier Data Comparison

One of the goals of this thesis is to determine whether or not the CNN is beneficial in classifying hyperspectral imaging data. To demonstrate this, the CNN classifier is compared to a KNN classifier and a logistic regression (LR) classifier, both using the SSH dataset. The KNN classifier makes predictions based on the K-nearest inputs trained on the classifier. For this classifier, K=5 is used as the number of nearest neighbors and Euclidean distance is used when classifying its results. The LR classifier uses the logistic function to determine the probability that each input belongs to the positive class (the nonlesional images, in this case).

The results of this comparison are shown in Table 9. The CNN outperforms the two simple classifiers, although the LR model is close to the CNN's performance, with 60.9% accuracy and a 0.566 F1 score.

Table 9

Simple Classifier Comparison

Models	Acceleration	Precision	Recall	F1 Score
KNN (K=5)	0.480	0.478	0.494	0.486
Logistic Regression	0.609	0.631	0.515	0.566
CNN	0.659	0.689	0.577	0.627

The CNN outperforms the KNN model by 17.9% in accuracy and by 0.141 in F1 score. The CNN outperforms the LR model by 5.0% in accuracy and by 0.061 in F1 score. The confusion matrix for the LR classifier is shown in Table 10. From this table, we see that the LR model is still classifying more lesional samples correctly compared to nonlesional samples.

Table 10
Confusion Matrix for the LR LOPO Validation Set

LR		
	Predicted Labels	
True Labels	Lesional	Nonlesional
Lesional	62	26
Nonlesional	42	45

6.3 Fine Tuning of CNN

In this section, the CNN fine tuning is presented and the results of the tuning are discussed. The SSH network is chosen as a baseline and tuned in the several ways. Models are evaluated on the basis of the activation function used, on the normalization method used, on the performance of 3-D convolution, on the number of convolution layers used, on the number of feature maps used, and on the number and size of the of fully connected layers used. Models using different activation functions, such as tanh,

relu, elu, and leaky relu are evaluated and the results of these evaluations are shown in Table 11. While the model using the elu activation function come the closest to matching the performance of the tanh activation function, it only performs as well as than the LR classifier from the previous set of models.

Table 11
Activation Function Tuning

Models	Acceleration	Precision	Recall	F1 Score
relu	0.546	0.550	0.486	0.515
leaky relu (α=0.3)	0.569	0.574	0.515	0.542
leaky relu (α=0.4)	0.560	0.564	0.509	0.535
leaky relu (α=0.5)	0.571	0.577	0.510	0.541
elu	0.615	0.636	0.534	0.579
tanh	0.659	0.689	0.577	0.627

Models using different methods of data normalizing are also examined. The baseline network normalizes the images based on the values of the camera output. Since this was a 14-bit camera, the maximum value of 16383 is used normalize each pixel to a value between 0 and 1. For these evaluations, we compare our original normalization method to a method normalizing each image based on the values of the image (min-max normalization) and to a method dividing each image by the maximum value for each image. The results of these evaluations are shown in Table 12.

Table 12

Normalization Tuning

Models	Acceleration	Precision	Recall	F1 Score
Image Max	0.525	0.525	0.468	0.494
Image Normalization	0.535	0.536	0.475	0.503
Dataset Normalization	0.659	0.689	0.577	0.627

Neither of the two image-based normalization methods perform well on the SSH dataset, with the Image Max model having an accuracy of 52.5% and the Image Normalization model having an accuracy of 53.6%.

The baseline model using 2-D convolution layers is compared to two models using 3-D convolution. The first model uses fewer feature maps in the second and third convolution-max pool layer pair. The second 3-D model keeps the same number of feature maps as the baseline model. The results of these evaluations are shown in Table 13.

Table 13

3-D Convolution Comparison

Models	Acceleration	Precision	Recall	F1 Score
2-D Convolution	0.659	0.689	0.577	0.627
3-D Convolution 1	0.507	0.508	0.288	0.367
3-D Convolution 2	0.488	0.483	0.408	0.436

Neither of the 3-D convolution layer models fare well, with the first 3-D convolution model having an accuracy of 50.7% and the second 3-D convolution model having an accuracy of 48.8%. More work will need to be done to properly evaluate the performance of the 3-D convolution layers.

The number of convolution layers for this classifier is also examined, with a comparison of the baseline model to models using 2 convolution-max pool layer pairs and 4 convolution-max pool layer pairs. The 4-layer model creates the extra pair by duplicating the third layer and the 2-layer model deletes the second pair of layers. The results of these evaluations are shown in Table 14.

Table 14

Convolution Layer Comparison

Models	Acceleration	Precision	Recall	F1 Score
2 Convolution and Max Pool Layers	0.569	0.586	0.453	0.509
3 Convolution and Max Pool Layers	0.659	0.689	0.577	0.627
4 Convolution and Max Pool Layers	0.578	0.577	0.566	0.571

Changing the number of convolution layers does little to help the classifier performance. The model built with two convolution-max pool layers has an accuracy of 56.9%. The model built with four convolution-max pool layers has an accuracy of 57.8%.

Models are also developed to determine the optimal number of feature maps, and the results of these evaluations are shown in Table 15. Tuning the feature maps does little to the CNN performance. The model with half the feature maps had a 64.2% accuracy, underperforming the baseline model by 1.7%. The model with twice the feature maps had a 61.2% accuracy, underperforming the baseline model by 4.7%.

Table 15
Feature Map Comparison

Models	Acceleration	Precision	Recall	F1 Score
1/2 Baseline	0.642	0.665	0.563	0.609
Baseline	0.659	0.689	0.577	0.627
2 $\times$ Baseline	0.612	0.630	0.539	0.580

The last set of evaluations compare the number fully connected layers (FC) at the output of the network. The model with two FC has 36 neurons in the first layer and 24 in the second. The model with the bigger FC has 36 neurons, while the model with the smaller FC has 12 neurons. The results of this evaluation are shown in Table 16.

Table 16

Fully Connected Layer Comparison

Models	Acceleration	Precision	Recall	F1 Score
Smaller FC	0.658	0.683	0.588	0.630
Baseline	0.659	0.689	0.577	0.627
Bigger FC	0.653	0.671	0.596	0.630
2 FC	0.645	0.667	0.574	0.615

The model with the smaller FC comes closest to the baseline model, with an accuracy of 65.8%, but has a slightly higher F1 score, with a score of 0.630. This model seems to be slightly more fair than the baseline model. The model with the bigger FC had an accuracy

of 65.3%, with an F1 score of 0.630. The model with the second FC layer has an accuracy of 64.5%, with an F1 score of 0.615.

None of these methods for fine tuning the model have done much to improve the model. Again, this is due to the model being at the limits of what can be done with the current small patient set. More data will be needed to build a better, more accurate network.

6.4 Further Discussion

This section looks at some of the spectra of the classified and misclassified images and takes a deeper look at the best and worst performing patient data from the classifier. Here are some of the spectra from the correctly classified and the misclassified images. Fig. 10 and Fig. 11 both show spectra from tissue samples that were correctly identified by the baseline SSH classifier.

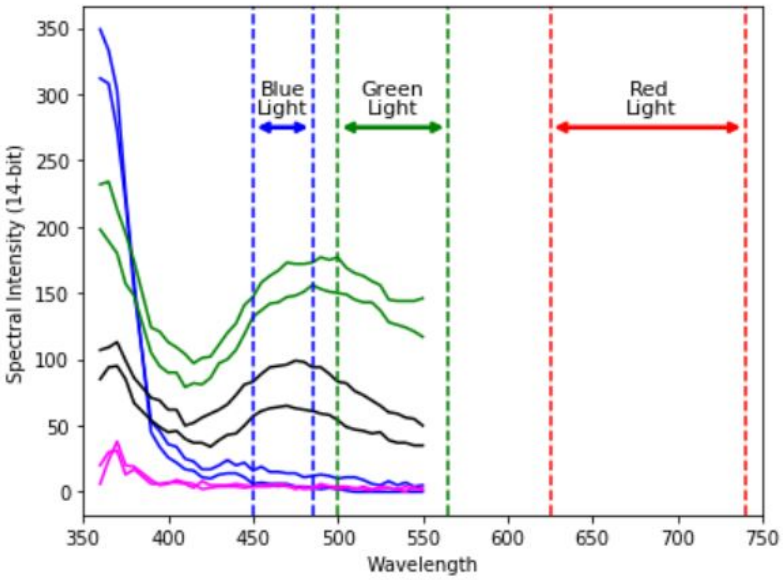

Fig. 10. Spectra from correctly identified lesional tissue samples.

Each color spectra represents spectra from the same patient. The dotted lines correspond to the wavelengths associated with blue, green, and red light. These two figures both

appear similar, except for a small spike in the blue wavelength range in the nonlesional spectra. We believe that this similarity may explain some of the difficulty in classifying hyperspectral data. This also demonstrates the challenge in processing hyperspectral data and why machine learning tools such as CNNs are needed to dig deeper into the data.

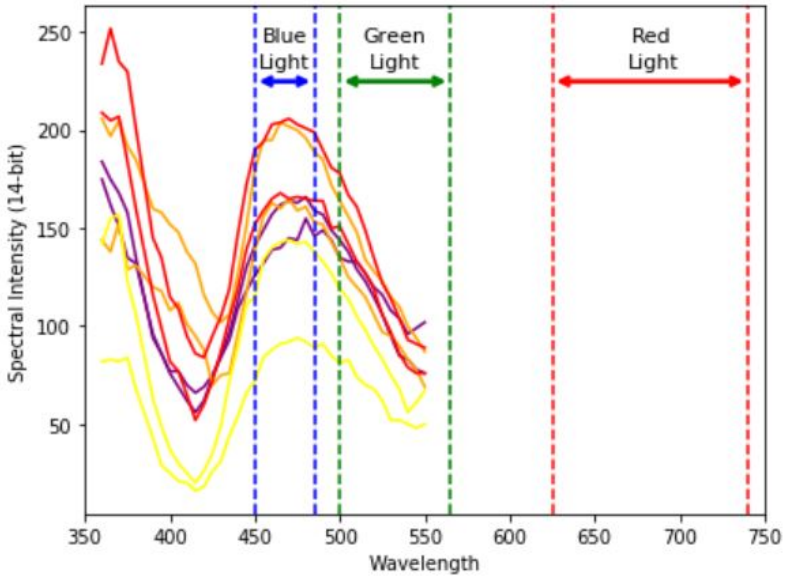

Fig. 11. Spectra from correctly identified nonlesional tissue samples.

We take a look at some of the errors in the network to better understand the CNN's performance. Fig. 12 and Fig. 13 both show spectra from tissue samples that were misclassified by the baseline SSH classifier. Again, each color spectra represents spectra from the same patient and the dotted lines correspond to the color of light for that wavelength. The spectra from the misclassified lesional images appear to share some of the same blue wavelength spike as the nonlesional spectra. The spectra misclassified as nonlesional appear to be flatter, except for two spectra that are much higher than the others. Still, it is hard to draw meaningful conclusions from these errors.

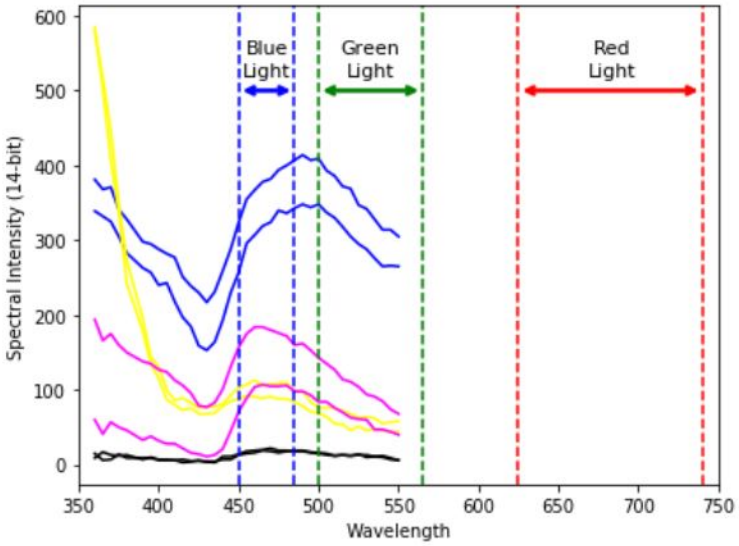

Fig. 12. Spectra from misidentified lesional tissue samples.

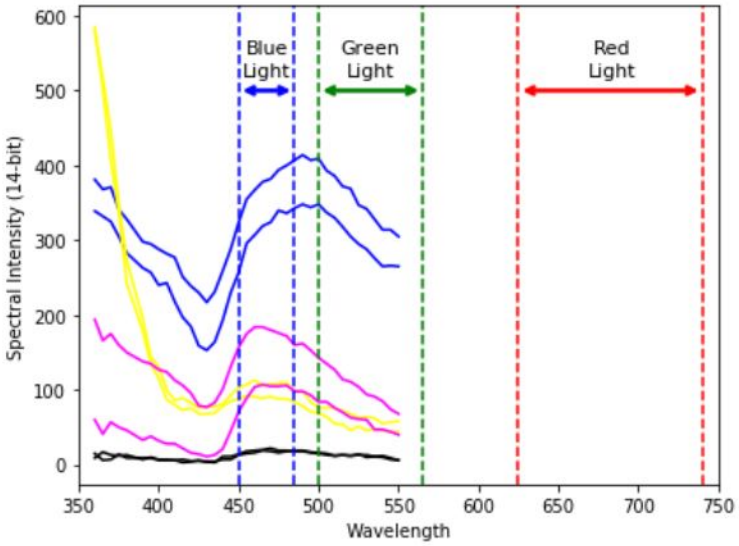

Fig. 13. Spectra from misidentified nonlesional tissue samples.

The model performs the worst against patient 3's data (Validation fold 8) when compared to each of the other patients' data. The model classifies every image segment of patient 3's data as lesional. To examine this, Fig. 14 looks at spectra from every

nonlesional image in patient 3's data. The spectra from the same image are represented as the same color. This figure shows spectra from each of the 5 nonlesional images from patient 3's data. Next this is compared to the lesional spectra from patient 3's data, shown in Fig. 15.

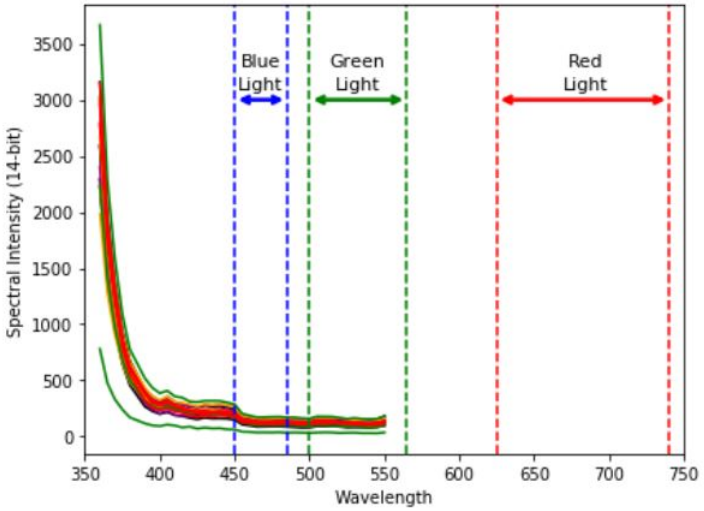

Fig. 14. Spectra from patient 3's nonlesional images (Validation fold 8).

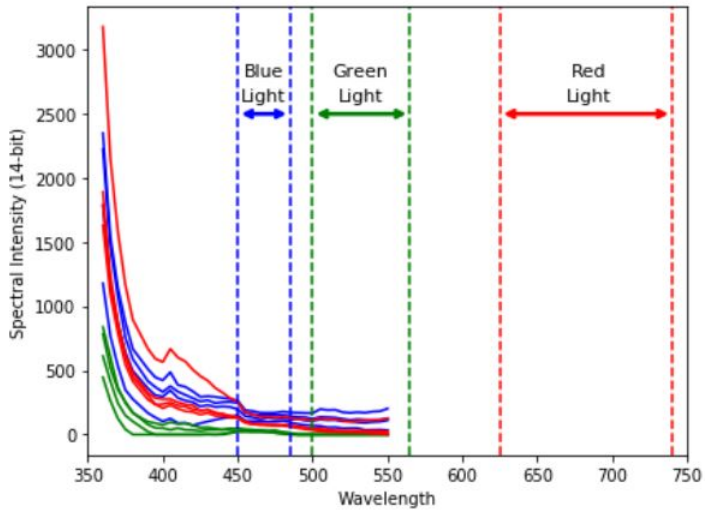

Fig. 15. Spectra from patient 3's lesional images (Validation fold 8).

The data for patient 3 have 3 lesional images, and spectra from each image is shown as a different color. There appears to be little differentiating the two sets, which may explain why they are all classified one way. The spectra from both these sets of data seem to have much higher spectral intensities in the 350 nm to 400 nm range than the other spectra shown in this section. It is possible that the data from patient 3 are an outlier and the extremely high spectral values in the ultraviolet portion of the spectrum are causing problems for the CNN.

Next we examine the data that fit the best to our model. The model performs best with the data for patient 6 compared to the data for the other patients. Not only does this fold have 88% accuracy, but its precision is 1, meaning no lesional samples are mistakenly classified as nonlesional. Fig. 16 shows spectra from the 6 nonlesional images in patient 6's data, color coded by the different images.

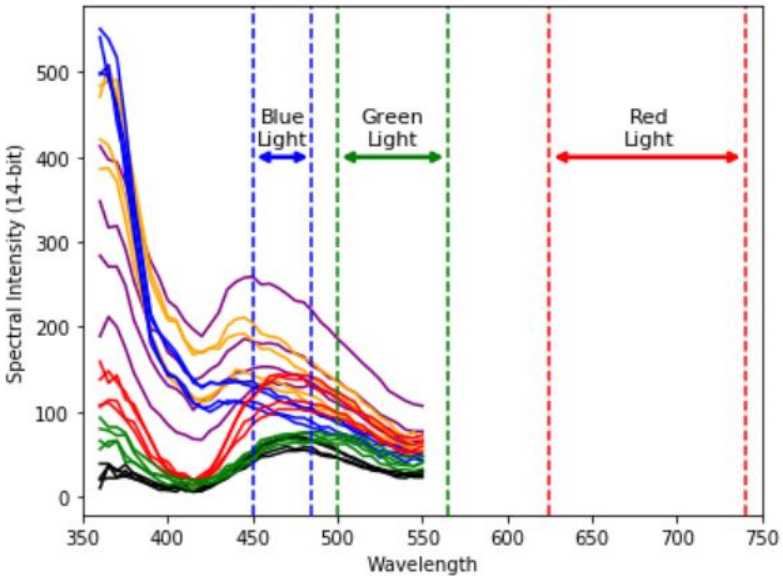

Fig. 16. Spectra from Patient 6 (Validation fold 9) nonlesional sample.

This is compared to Fig. 17, which shows spectra from the 4 lesional images in patient 6's data.

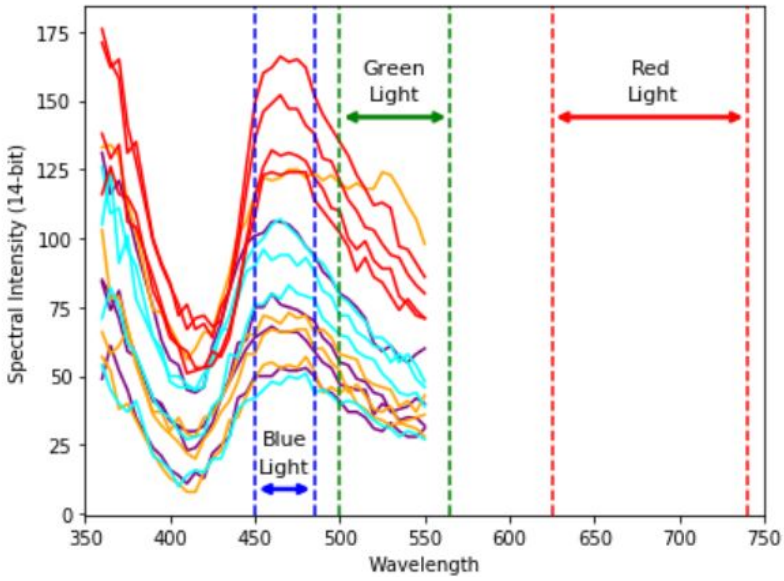

Fig. 17. Spectra from Patient 6 (Validation fold 9) lesional sample.

Both of these sets of images for patient 6's data contain much lower spectral intensities than the images for patient 3's data. These are closer to the typical spectra that were shown before. The images show that the cancerous images have a higher spike in the blue portion of the spectrum. However, the lesional images from patient 6's data seem to have lower intensities overall.

7 FUTURE WORK

One of the main problems with this dataset is the small number of patients. The two main paths of future work involve ways that could improve the usefulness of the CNN classifier if a larger pool of patient images becomes available, and the paths that could improve the usefulness of the model if the number of patients remains small.

If more patients can be studied, then the first major area of exploration would involve developing 3-D convolution layers to better tease out the differences in the spectra. Current exploration into these convolutional networks did not yield much success, but further study would be required to know whether or not they can improve the classification process. There is reason to believe with the importance of the spectral information that the 3-D kernel may ultimately be better attuned to capturing these changes. The second major area of exploration would be the use of PCAs with the hyperspectral images. PCAs are often useful in indicating which features contribute most to an output, and they could be useful for indicating which bands tend be more active in the cancerous cells. PCAs could also be used in tandem with CNNs in order to reduce the number of features of a hyperspectral image. In addition to these ideas, further data augmentation might be feasible. While the current use of image segmentation may have been pushed to its limits, with an increased patient pool, other data augmentation methods may be worth exploring.

If the number of patients in the hyperspectral dataset remains small, a different set of paths will need to be explored. In the case of small datasets like the one available for this thesis, transfer learning can be a useful tool for building new models. While most of the publicly available datasets involve typical 3-channel images, if used in conjunction with other machine learning tools like PCAs, the hyperspectral images might be applicable to those networks as well. There may be possible ways to combine the hyperspectral information with information found in common grayscale colonoscopy images and other

common medical tools. These could be used with PCAs and other current hyperspectral fusion methods to build up a better network for the classifier with the limited information available.

8 CONCLUSIONS

This thesis demonstrated that information from hyperspectral data were able to increase the performance of the developed colon cancer sample classifier. It was shown that the SSH dataset had a 5.6% improvement in accuracy and a 0.037 F1 score improvement over models using grayscale images. It also showed how CNNs were useful in better classifying the hyperspectral data, as the CNN showed a 5% accuracy improvement and a 0.061 F1 score improvement over the simpler classifiers such as KNN and LR classifiers.

One of the factors limiting the performance of this system is the current number of patients available for this study. With the high variability between patients, one the most important things to do to improve the system would be to get more patients into the study, either through more hyperspectral studies or through data fusion with current colonoscopy methods.

Once further researched, hyperspectral imaging could become a critical tool in catching colon cancer in its earlier stages. In the future, understanding the spectral makeup of these tumorous cells could make it easier to detect cancer before it grows to the point that it becomes life-threatening.

Literature Cited

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2019," *CA: A Cancer Journal for Clinicians*, vol. 69, pp. 7–34, Jan. 2019.

[2] R. L. Siegel, K. D. Miller, and A. Jemal, "Colorectal cancer statistics, 2017," *CA: A Cancer Journal for Clinicians*, vol. 67, pp. 177–193, March 2017.

[3] S. Rathore, M. Hussain, A. Ali, and A. Khan, "A recent survey on colon cancer detection techniques," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 10, pp. 545–563, July 2013.

[4] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geoscience and Remote Sensing Magazine*, vol. 1, pp. 6–36, July 2013.

[5] W. Mcculloch and W. Pitts, "A logical calculus of ideas immanent in nervous activity," *Bulletin of Mathematical Biophysics*, vol. 5, pp. 127–147, 1943.

[6] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[7] A. Gron, *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, Inc., 1st ed., 2017.

[8] M. J. Khan, H. S. Khan, A. Yousaf, K. Khurshid, and A. Abbas, "Modern trends in hyperspectral image analysis: A review," *IEEE Access*, vol. 6, pp. 14118–14129, 2018.

[9] J. M. Bioucas-Dias and A. Plaza, "An overview on hyperspectral unmixing: Geometrical, statistical, and sparse regression based approaches," in *2011 IEEE International Geoscience and Remote Sensing Symposium*, pp. 1135–1138, July 2011.

[10] G. Lu and B. Fei, "Medical hyperspectral imaging: A review," *Journal of Biomedical Optics*, vol. 19, pp. 1 – 24, Jan. 2014.

[11] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, pp. 2278–2324, Nov 1998.

[12] D. Bazazeh and R. Shubair, "Comparative study of machine learning algorithms for breast cancer detection and diagnosis," in *2016 5th International Conference on Electronic Devices, Systems and Applications (ICEDSA)*, pp. 1–4, Dec 2016.

[13] J. Alam, S. Alam, and A. Hossan, "Multi-stage lung cancer detection and prediction using multi-class svm classifie," in *2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2)*, pp. 1–4, Feb 2018.

[14] Y. J. Tan, K. S. Sim, and F. F. Ting, "Breast cancer detection using convolutional neural networks for mammogram imaging system," in *2017 International Conference on Robotics, Automation and Sciences (ICORAS)*, pp. 1–5, Nov 2017.

[15] A. M. Godkhindi and R. M. Gowda, "Automated detection of polyps in ct colonography images using deep learning algorithms in colon cancer diagnosis," in *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, pp. 1722–1728, Aug 2017.

[16] T. Chiang, Y. Huang, R. Chen, C. Huang, and R. Chang, "Tumor detection in automated breast ultrasound using 3-d cnn and prioritized candidate aggregation," *IEEE Transactions on Medical Imaging*, vol. 38, pp. 240–249, Jan 2019.

[17] S. J. Leavesley, M. Walters, C. Lopez, T. Baker, P. F. Favreau, T. C. Rich, P. F. Rider, and C. W. Boudreaux, "Hyperspectral imaging fluorescence excitation scanning for colon cancer detection," *Journal of Biomedical Optics*, vol. 21, pp. 1 – 10, Oct. 2016.

[18] G. Lu, L. V. Halig, D. Wang, X. Qin, Z. G. Chen, and B. Fei, "Spectral-spatial classification for noninvasive cancer detection using hyperspectral imaging," *Journal of Biomedical Optics*, vol. 19, pp. 1–18, Oct. 2014.

[19] M. Nathan, A. S. Kabatznik, and A. Mahmood, "Hyperspectral imaging for cancer detection and classification," in *2018 3rd Biennial South African Biomedical Engineering Conference (SAIBMEC)*, pp. 1–4, April 2018.

[20] R. Pike, G. Lu, D. Wang, Z. G. Chen, and B. Fei, "A minimum spanning forest-based method for noninvasive cancer detection with hyperspectral imaging," *IEEE Transactions on Biomedical Engineering*, vol. 63, pp. 653–663, March 2016.

[21] A. Gopi and C. S. Reshmi, "A noninvasive cancer detection using hyperspectral images," in *2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, pp. 2051–2055, March 2017.

[22] S. J. Leavesley, J. Deal, S. Hill, W. A. Martin, M. Lall, C. Lopez, P. F. Rider, and C. W. B. Thomas C. Rich, "Colorectal cancer detection by hyperspectral imaging using fluorescence excitation scanning," in *SPIE BiOS, 2018*, vol. 10489, Feb. 2018.

[23] L. Ma, G. Lu, D. Wang, X. Wang, Z. G. Chen, S. Muller, A. Chen, and B. Fei, "Deep learning based classification for head and neck cancer detection with hyperspectral imaging in an animal model," in *SPIE Medical Imaging*, vol. 10137, March 2017.

[24] J. Deal, S. Mayes, C. Browning, S. Hill, P. Rider, C. Boudreaux, T. C. Rich, and S. J. Leavesley, "Identifying molecular contributors to autofluorescence of neoplastic and normal colon sections using excitation-scanning hyperspectral imaging," *Journal of Biomedical Optics*, vol. 24, no. 2, pp. 1 – 11, 2018.