2018

Use of Hyperspectral Images (HSI) and Convolutional Neural Network (CNN) To Identify Normal, Precancerous and Cancerous Tissues

Pallavi Jain
Technological University Dublin, Ireland

Use of Hyperspectral Images (HSI) and Convolutional Neural Network (CNN) To Identify Normal, Precancerous and Cancerous Tissues.

Pallavi Jain

A dissertation submitted in partial fulfillment of the requirements of Dublin Institute of Technology for the degree of M.Sc. in Computing (Data Analytics)

September 2018

DECLARATION

I certify that this dissertation which I now submit for examination for the award of MSc in Computing (Data Analytics), is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

This dissertation was prepared according to the regulations for postgraduate study of the Dublin Institute of Technology and has not been submitted in whole or part for an award in any other Institute or University.

The work reported on in this dissertation conforms to the principles and requirements of the Institutes guidelines for ethics in research.

Signed: ----------

Pallavi Jain

Date: 03 September 2018

Abstract

Cancer detection has been a great topic of research for a long time, as early detection of cancer can help in increasing the survival rate of patients by providing on time better treatment. A robust system is required in order to detect early-stage cancer as its difficult to identify early-stage cancer from the normal clinical process. The computer vision techniques provide a new way to understand the challenges related to the medical image analysis.

This thesis presents the medical image analysis using a combination of Convolutional Neural Network and Hyperspectral Images of cancer patient's tissues. The idea behind choosing the CNN is it has been doing really well in image processing and outperformed the other traditional techniques. An attempt is made to distinguish between Normal Tissues, Premature Tissues and Oesophageal adenocarcinoma (OAC) tissues. The dataset used here posses many challenges like less number of instances and most importantly imbalanced data, which means some instances are very few in comparison to others. This thesis focuses on improving the F1 Score of the CNN classifier and the performance is measured after fine-tuning the baseline model. The experiment result shows that fine-tuning the CNN algorithm help in improving the F1 Score a bit though haven't achieved great result due to the limitation of imbalanced data. This work is a contribution towards detection of early-stage cancer through images, which clinical processes are unable to detect.

Keywords: *Convolutional Neural Network, Hyperspectral Images, Image Processing*

Acknowledgement

In the process of this thesis, I'd like to thank my supervisor Dr. Robert Ross for his consistent support and encouragement while doing this project. His in-depth knowledge and expertise in this subject helped a lot, apart from that his ease of working and providing proper working schedule helped a lot to complete the thesis on time.

I am also appreciative of Dr. Thi Nguyet Que Nguyen for her continuous support and guidance throughout the project, providing all the details regarding the data and helped in strategy planning and her encouragement to me in order to pursue this topic.

I'd like to thanks Dr. Aidan Meade, for providing the data while remaining considering all the data protection issues, he also helped in providing information regarding the creation of data and provided all necessary help during the thesis.

Table of Contents

List of Figures

List of Tables

CHAPTER 1

1. Introduction

Oesophageal adenocarcinoma (OAC) is one of the deadliest cancer and in recent years its incidence increased in fast pace, its 5-year survival in the United States ranges from 14% to 22%. It usually gets detected in advanced stages where only 60% of patients are suitable for palliative therapy (Besharat et. al., 2008). In the process of development of OAC, there are many intermediate stages, which starts from a premalignant lesion that is Barrett Oesophagus (BO), the histological changes start from BO to low-grade dysplasia (LGD) to high-grade dysplasia (HGD) and to OAC (Feber et al., 2008). The early diagnosis of BO can not decrease the mortality rate, but it can let the patient know the endoscopy surveillance requirements and what treatment like chemoradiation or surgery can help the patient for survival. But as most of the intermediate stages looks like normal tissue it is difficult to diagnose with normal clinical processes (Corley et al.2002). Also according to Shiozaki et. al., OAC has a character that is emphasized by the development of hematogenous and lymphatic metastases from submucosal lesions that may not show any symptoms and may be difficult to detect by barium study. Due to this, oesophagal cancer patients have poor survival rates despite having the option of radical surgical treatment or radiation therapy. This leads us to the need for early detection before a tumour attacks the submucosal layer (Shiozaki et. al.,1990). Because of all the above reason diagnosing the OAC and its intermediate stages has gained big time attention of researchers. Though there is not much research evidence on the diagnosis of OAC using deep learning and machine learning. This research is conducted in order to see whether a *deep learning algorithm can distinguish between Normal tissues, intermediate and OAC stages tissues using hyperspectral images.*

1.1 Background

Cancer diagnosis through the clinical process is still very complex and uncertain process; which directs us to automate the process for better accuracy. Traditional machine learning processes are efficient in detecting cancer cells but require domain-specific expert knowledge, due to high dimensionality in gene expression data, which requires feature extraction for better performance in detection (Danaee et.al., 2017).

Medical image analysis is nowadays are highly helpful in disease detection and the great interest of research. In a medical application, hyperspectral imaging is the emerging technique, which can be used to detect cancer noninvasively. Hyperspectral Images (HSI) are quite efficient in the diagnosis of cancer as it is not limited to RGB features of images but it also consists of spectral and spatial features. In HSI, tissue absorbs many wavelengths of infrared or near-infrared and provide the images with pixels where each pixel has multiple wavelengths and spatially spreaded. HSI can be of a great source of information for detection, as spectral information tells the characterization and classification of the different type of tissues whereas spatial information provides the difference in reflectance of cancer and normal tissue (Akbari et al., 2012).

The information given by HSI can be used to automatically identify cancerous and normal tissue region using machine learning and deep learning. This segmentation can also help during surgery where real-time information can be provided. Deep learning technique especially convolutional neural network (CNN) has been performing really well and outperforming in visual image classification like object detection, and image segmentation (Hu et al., 2015).

This research mainly focusing on Oesophagus Cancer detection along with normal and intermediate stages of cancer for early detection using deep learning technique CNN and hyperspectral images.

1.2 Research Problem

Although traditional machine learning classification models like SVM, KNN performs really well on the classification of cancer cells using hyperspectral images. However, these technologies have limitations, as it requires domain-specific knowledge to extract features to convert suitable features to feature vector representation. And cancer cell information comes with the high dimensionality due to this, their performance becomes highly dependent on the feature extraction process (Yang et.al., 2016).

Also, in normal fluorescent microscopy imaging, many dyes have an overlapping of emission spectra due to which it is not able to identify dyes, which has a similar spectral emission (Schultz et al., 2001). Due to this limitation of technique, most of the important information about the genes stay hidden and ultimately efficiency of

classifiers to identify cancer cells reduces.

The CNN alleviates the problem of feature extraction due to its self- deep learning process (Yang et. al., 2016) whereas hyperspectral images gives spatial and spectral dimensions which can help to retrieve more information with images for disease detection (Lu & Fei, 2014), and ultimately raises the research question that is:

"*Does Convolutional Neural Network algorithm strong enough to identify normal, intermediate and cancer cells using hyperspectral images*?*"*

1.3 Research Objective

Succinctly, the aim of this research is to evaluate the power of CNN algorithm and hyperspectral images together to distinguish between the cancer cells along with the normal and intermediate cells. In this regard, and in light of the existing literature explored below in Chapter 2, the effective Null Hypothesis is that the CNN and hyperspectral images provide no predictive power for future cancer diagnosis or real-time surgery. While the hypothesis for this project is to see whether CNN and hyperspectral images together provide a powerful model to identify cancerous, normal and intermediate cells from images. The project has the baseline model which is a basic CNN architecture shown in Chapter 4, on this baseline model the research is carried out which will show if there is any improvement on the prediction of different types of cells with the changes in hyperparameters and different sampling techniques. To achieve the results in identifying a different kind of cells, the research carried out in a process, which is outlined below:

1. Exploring the previous works on identifying the cancer cells and providing the comprehensive analysis, to carry out the further research and providing the best predictive model with deep learning.
2. Data processing will be done which includes image resizing, converting raw images to the 2nd derivative image, normalization.
3. Different sampling techniques will be used for the continuous improvement of the basic model.
4. Build the CNN model to increase the prediction power in order to identify different types of cells.
5. Comparing the evaluation results of a different version of the model and selecting the best among them.

1.4 Research Methodologies

The focus of this research is to get the comparison of a different version of CNN model architecture with different sampling techniques on the existing dataset; therefore it comes under the secondary research. As part of this existing research, a literature review was carried out for the image classification algorithms, and image preparation to get the comprehensive idea of the project.

The research work follows the quantitative (Epidemiological) methodology and it is empirical in nature. The experiment is carried out to get the results for different parameters and sampling technique, which will then verify the given hypothesis. Experiment results are then evaluated to check the performance of the classifier by comparing the F1 Score of a different version of the model. It is an inductive experiment as the result is based on the model with the best F1 score, and showing that already existing CNN or deep learning algorithm can be used to identify cancer cells.

1.5 Scope and Limitations

The scope of the experiment is limited to one type of cancer that is Oesophagus adenocarcinoma. The experiment is referring to the early stages of cancer along with the normal and final stage of cancer tissues.

The research carried out on 56 patient's tissues whose tissue has been recorded and observed using FTIR imaging which created 85 instances combined of Normal, Intermediate and Cancer tissues. The data basically consist of hypercube with a matrix of pixels and spectral depth.

The focus of this research is limited to basic CNN model architecture with variation in the data sampling and processing techniques along with hyperparameter variations to evaluate the best possible way to identify cancer in the dataset.

The limitation to this research is the size of data, and unbalanced classes of tissues as some tissues has very less number of instances which makes it difficult to feed enough images to train the classifier for all the class of images. To provide the more appropriate results in future, more number of images required identifying the particular type of cancer.

1.6 Dissertation Outline

- Chapter 1 (Introduction) was devoted to the introduction of image classification and background about oesophageal cancer and its intermediate stages. It also

covers the problem statements being solved along with an overview of related work in the image classification problem. It describes the research methodology, scope and limitation of research.

- Chapter 2 (Literature Review) highlights the concept of state-of-the-art related computer vision problem for the purpose of filling gaps in the research to propose the research question for this thesis.

- Chapter 3 (Design and Methodology) will explain about the design plan of the experiment along with Dataset details, Model Architecture and Evaluation metrics.
- Chapter 4 (Implementation and Results) gives the details about the convolutional neural network implementation and associated result with each version of the model.
- Chapter 5 (Evaluation/Analysis) will give the detailed analysis of the experiment and based on the result a decision regarding the acceptance and rejection of the proposed hypothesis will be made. This chapters also outlines the strength and weakness of the project.
- Chapter 6 (Conclusion) will summarize the working and finding of the research undertaken during this thesis work, which includes problem definition, network design, experiment setup and evaluation of the finding and limitation for further work.

CHAPTER 2

2. LITERATURE REVIEW

The chapter provides a detailed review of relevant literature about the application of computer vision in the field of cancer detection. Previous works on medical image analysis using machine learning techniques and deep learning techniques has been mentioned along with the outcome and shortcomings of those techniques. Apart from the previous state of the art, there is a review of some of the data handling techniques, which will be used in this thesis to improve the performance. The review is broadly classified into two sections, Medical image analysis and Deep learning.

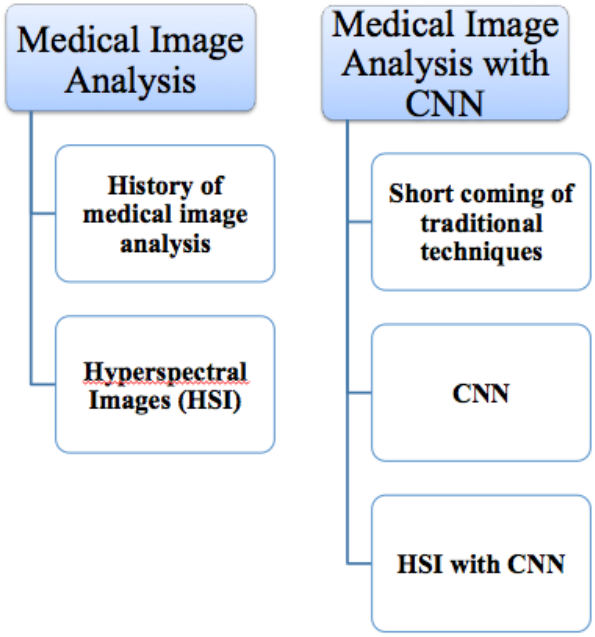


Figure 2.1 Literature Review Structure

2.1 Medical Imaging

2.1.1 Medical Image Analysis

The medical image analysis has evolved a lot in last 20 plus years, earlier there was no term as medical image analysis but over the time it came into the medical terms and changed the diagnosis methods of the medical field. Most of the work in medical image analysis is motivated for pattern recognition, image processing and computer

vision. Earlier an only a small number of researchers used to think that medical image analysis is unique information and which can help in image processing and pattern detection (Duncan & Ayache, 2000). One such kind of work has been done by Ballard & Sklansky who located tumours using pattern recognition methods (Ballard & Sklansky, 1973). Later in 1987, computer-aided diagnosis of x-ray mammograms was proposed by Chan et al, in which they did feature detection to classify normal tissue and calcification (Chan et al., 1987). By the time in the 1990s, image data availability and quality improved a lot like 3D Magnetic resonance images (MRI), spiral-CT, full 3D ultrasound, and the more routine use of 3D SPECT and PET. With 3D MRI images the spatial information came into the picture which helped in giving better visualisation and accurate segmentation (Andreasen et al., 1994). Some of the disease characteristics can only be detected by histopathology images. The development of these images preserves the original tissue structure due to which its gold standard which helps in identifying any type of cancer can also be preserved. Apart from that it also gives the spatial information about the nuclei in the tissue. (Gurcan et al., 2009).

2.1.2 Hyperspectral Imaging (HSI)

Later more type of medical images was developed which provided spectral and spatial information of the tissues which could help in finding more features for disease detection, and one of them is multispectral or hyperspectral images which can be seen in. Originally HSI developed was for geologic surveys, which has been now highly adopted as an effective and noninvasive tool in the clinical diagnosis of medical image analysis (Siddiqi et al., 2008)

A paper by Schultz et al showed that current fluorescent microscopy images have a limitation as many similar dyes have overlapping emission spectra. Due to this limitation of technique, most of the important information about the genes stay hidden and ultimately efficiency of classifiers to identify cancer cells reduces. So to capture the whole spectrum emission they proposed the hyperspectral images, which has the whole spectrum for each pixel of the image (Schultz et. al., 2000).

As an application of the HSI, the paper proposed by Lu and Fei on Hyperspectral Images in the medical field, they found that HSI could be successfully implemented in the non-invasive disease diagnosis and surgical guidance. They showed that HSI consists of both spatial and spectral domain, which can help to retrieve more information with images for disease detection (Lu & Fei, 2014).

Spectral measurement and analysis of the prostate can be done without physical contact and can be performed without extracting any tissue. Apart from that many tissues can be measured simultaneously. HSI could potentially be used to detect residual tumours by surveying and noninvasively examining a vast tissue area and without the need to obtain tissue samples (Akbari et.al., 2012).

Martin et al performed the experiment of creating hyperspectral images of a tumour in a mouse and they were able to detect the normal as well as malignant tissue using the difference between the wavelength intensities (Martin et al., 2006).

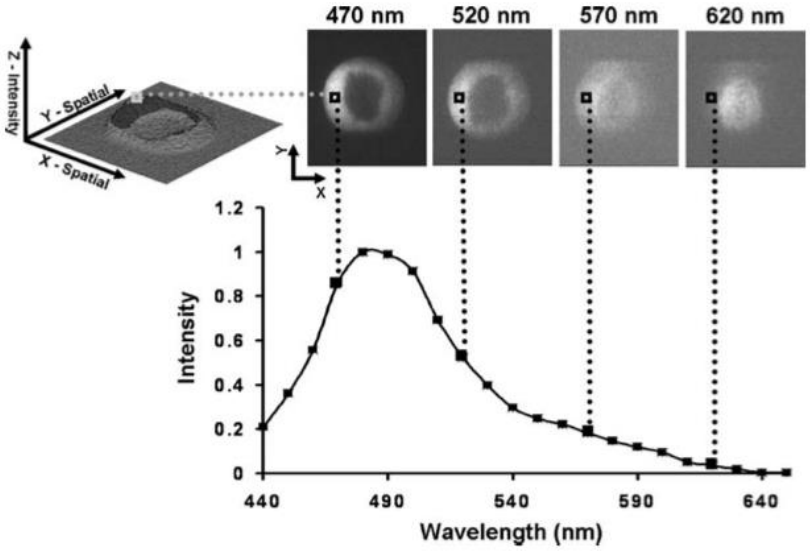


Figure 2.2 Hyperspectral Image Explanation (Martin et al.,2006)

2.2 Medical Image Analysis with CNN

2.2.1 Traditional Techniques

As the medical imaging has become an important technique to identify any disease or cancer tissues, many researchers started to find several machine learning and deep learning technique to detect cancer tissues accurately and precisely. In the light of cancer identification, machine learning came into picture, traditionally the identification has been done by gene expression data, it was an efficient approach for detecting cancer but required great domain knowledge to use for right minimal feature extraction to avoid overfitting of the data, as this data has high dimensionality due to which its difficult to extract features correctly (Danaee et. al., 2017).

In recent years, many advancements have been done in cancer cell identification with image processing data or gene expression data using several machine learning algorithms like SVM, KNN and Artificial Neural network. Unsupervised learning is also one of the good technique to divide the different type of data into clusters based on different features with similar patterns, but with gene expression data there can be many different relationships among genes which make it difficult for biological representation of data (Ramaswamy, & Golub., 2002).

More precisely in the machine learning world, Support Vector Machine (SVM) is one of the most popular techniques, which helps in overcoming the drawback of unsupervised learning of hierarchical clustering or self-organizing maps. SVM also has the ability to deal with high dimensional data of microarray gene expression as it allows to it cluster the genes in similar fashion in advance which are then combined to make training dataset (Brown et. al., 2000).

Although traditional machine learning classification models like SVM, KNN performs really well on the classification of cancer cells using images, however, these technologies have limitations, as it requires domain-specific knowledge to extract features to convert suitable features to feature vector representation. And cancer cell information comes with the high dimensionality due to this their performance becomes highly dependent on the feature extraction process (Yang et. al., 2016). In contrast to the conventional pattern recognition techniques, deep learning models can learn features by building high-level features from low-level ones, which ultimately reduce the steps for feature extraction (Makantasis et. al., 2015).

2.2.2 CNN

CNN has been used for a long time, LeCun has been working on image recognition since 1989, where he used backpropagation on handwritten digit recognition, later in 1995 he proposed a paper on CNN where he proposed the use of a fully connected multilayer network to classify images and to eliminate feature extraction (LeCun et al, 1989) and (LeCun et al, 1995). During this time CNN didn't get much popularity (Litjens et al., 2017), CNN gained popularity when Krizhevsky et al proposed AlexNet, in the ImageNet competition (Krizhevsky et al., 2012).

According to Krizhevsky et al, Convolutional Neural Network (CNN) is one of the deep learning techniques, which has a multilayer architecture and has good learning capacity for millions of images, in which controlling the depth and breadth can vary its

learning capacity. They have very fewer connections and parameters and so they are easier to train (Krizhevsky et. al., 2012).

As feature extraction is one of the biggest problems for better performance of SVM due to the high dimensionality of the data. There are many feature extraction techniques like Principle Component Analysis (PCA), which can reduce the features. Though according to Yang et al, CNN classifier can alleviate this problem, as it does not require feature reduction steps, and its training process itself consists of feature extraction (Yang et. al., 2016).

Many models have been proposed so far for image classification, in light of which paper proposed by Hu et al, According to him, CNN outperformed many traditional techniques in visual image classification like object detection and house number image. He proposed a model, which identifies images in the spectral domain. His architecture consists of a basic model as in figure 2.3 with a convolutional layer, pooling layer and fully connected layer (Hu et. al., 2015).

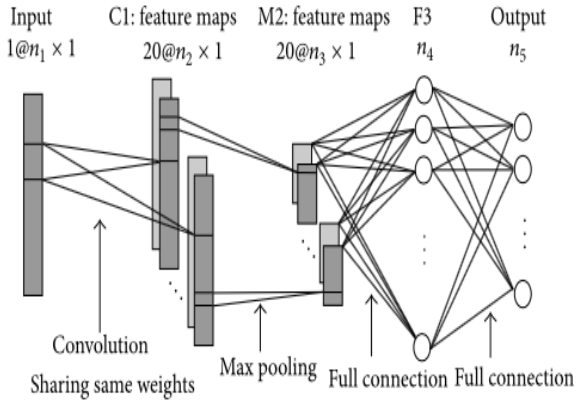


Figure 2.3 Hu et. al., proposed CNN architecture

2.2.3 Cancer detection with CNN and HSI

Though the above framework only consists of the spectral domain and no spatial domain. But CNN can be used for both spatial and spectral domain of images, the model proposed by Wang et. al., for breast cancer detection, in which they cascaded a CNN model and features like morphology, colour, and texture features in order to improve the performance of the detection of cancer (Wang et. al., 2014). Whereas Sirinukunwattana et.al., proposed a spatially constrained CNN (SC-CNN) approach, which trains to predict the probability of a pixel being the centre of a nucleus, which actually identifies the location of nuclei to detect cancer (Sirinukunwattana et.

al.,2016).Traditional neural network not shown as good performance as SVM, they were slow and too much dependent on data size for their performance (Hu et. al., 2015).

Halicek et al proposed a research on head and neck cancer using CNN and HIS and, his results show that CNN has potential to classify tissues using hyperspectral images and can be used as near real-time tissue labelling for surgeries. They also found that CNN outperformed all the other models with the highest accuracy of 96.4% as compared to all other algorithms like SVM or KNN, also he showed that CNN is much faster at prediction as it does not require any post-processing to get better results (Halicek et. al., 2017).

Mei et al, also proposed that the use of integrated information from spectral and spatial information leads to a significant improvement in the accuracy of classification. Also, they performed an experiment to distinguish ground objects using Hyperspectral images in both supervised and unsupervised mode with CNN and found that supervised mode outperformed as compare to unsupervised mode (Mei et. al., 2017).

2.3 Handling Imbalanced Data

The performance of an algorithm is dependent on many factors; one such factor is the data quality. The biggest problem with medical image analysis is having highly skewed data, where some type of instances is very low as compared to other instances. For the cancer data this problem is prevailing, cancer is the rare instance as compare to normal instances. This kind of problem is called as Imbalanced data and there is always a huge cost associated with the wrong prediction for those rare instances, like a patient with cancer, predicted as no cancer can be very devastated and costly for life. They will never know that they can have proper treatment or surgery on time to survive. For such scenario, it is necessary to handle the imbalanced data to get the better predictions. There are few ways to handle the effect of imbalance data, which are either handling it at data level or at algorithm level. Sampling techniques are the process of handling the effect at the data level. The sampling techniques to handle data are mostly of three types; those are minority oversampling, majority undersampling or combination of both. For doing oversampling you can increase data by duplicating instances but this has the drawback of overfitting of the model. One of the technique for oversampling has been proposed by Chawla et. al. that is Synthetic Minority Over-Sampling (SMOTE), which creates the synthetic samples of those classes which are

minor, equals to the majority class. The SMOTE follows the principle of minority nearest neighbour and creates the nearest sample to a particular category, which means it works with the feature space rather than data space (Chawla et al.,2002). The other method to handle the imbalance effect is cost sensitivity, like regularization, which highly penalizes the wrong predictions. Dropout is one of the recent regularization technique, it drops some of the networks temporarily, by which it minimizes the loss function stochastically while preventing from overfitting (Shrivastava et al.,2014).

2.4 Summary, Limitations and Gaps Of Literature

A detailed review of the application of deep learning algorithm in computer vision is studied as part of solving the problem of cancer detection in the field of medical image analysis. Initially, relevant literature about the evolution of the medical image analysis has been provided which showed that work on this field has been done since a long time and how the use of hyperspectral images came into the picture as a solution for medical image analysis.

In the next section, the emphasis is given to the image analysis, where image detection with traditional techniques like SVM and KNN has been reviewed along with the shortcoming of these techniques. This section also describes the Convolutional Neural Network and its benefits for solving the problem of traditional techniques. Later on this section, the brief review of the use of Hyperspectral images with the convolutional neural network has been provided, which shows how the use of this has improved the performance for other types of cancer.

It has been seen that in the field of cancer detection using the combination of HSI and CNN, not many research evidence has been found so far and is still the matter of research. There are many evidence of better performance in cancer cell detection with CNN with normal images or MRI, CT Scan images but still, with hyperspectral images, there are very few of them. As per the Halicek et al, the data limitation is the problem for HSI data, also use of proper sampling techniques, regularization techniques and the optimizer can optimize classification performance and improve generalizability to make a robust classifier (Halicek et. al., 2017). Also, his evidence is limited to the detection of normal and cancerous tissues, while malignant or premature cancer tissue needs the high attention in order to provide early treatment of patients, which can increase the survival rate.

The limitation and research gaps presented in this section can be addressed by the research question given as

"*Does Convolutional Neural Network algorithm strong enough to identify normal, intermediate and cancer cells using hyperspectral images*?*".*

The next sections will describe the research design, implementation and evaluation of experiment to address the research question

CHAPTER 3

3. Design and Methodology

In this Chapter, a detailed overview of the plan and design of the experiment has been elaborated, which will help in understanding the basis of an experiment in a better way. The experiment will follow the basic approach that is Data preparation, Model Architecture and Evaluation criteria. Using python programming language has done the implementation of the experiment, python has great deep learning library such as TensorFlow. TensorFlow is an open source deep learning library for high-performance numerical computation. This has been the most popular and powerful tool for the implementation of CNN due to its flexible libraries.

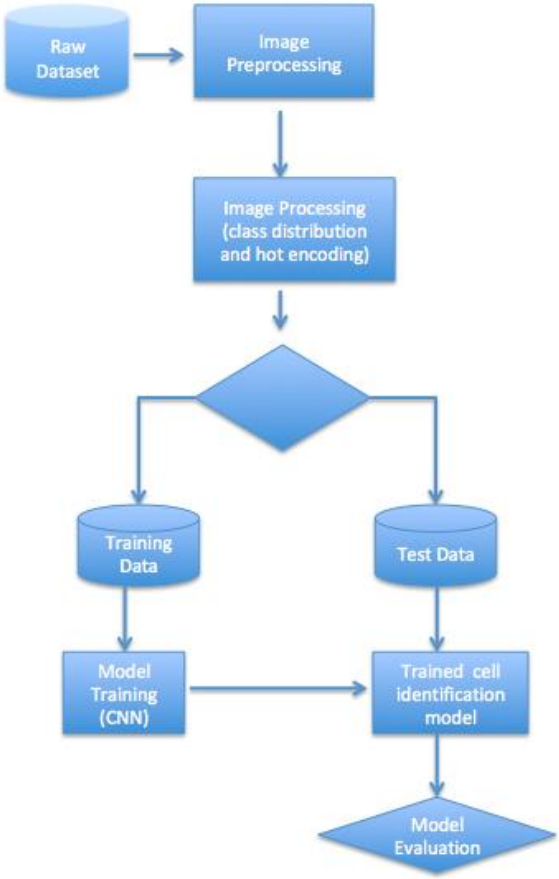


Figure 3.1 Experiment Design

The aim of this thesis to build a model which can distinguish between the normal tissues, intermediate tissues and cancerous tissues accurately and precisely. Figure 3.1 provides the high-level design of the experiment in order to build the model.

3.1 Business Understanding

The focus of this research is to improve the performance of basic CNN model with hyperspectral images in order to detect Oesophageal cancer accurately along with the normal and intermediate stages. In order to improve the performance, baseline CNN model has been built which will work as the benchmark and then it has been fine-tuned by using several techniques to get better F1 Score. The null hypothesis for this research can be stated as

H0: "Fine-tuning the basic CNN model do not provide a significant improvement in the classification of the normal, precancerous and cancerous in terms of the F1 score"

whereas research alternative hypothesis can be stated as

HA: "Fine-tuning the basic CNN model provides the significant improvement in the classification of the normal, precancerous and cancerous in terms of the F1 score"

3.2 Data Understanding

3.2.1 Dataset

Tissue Micro-Array (TMA) slide contains > 80 tissue of Barrett's Esophagus patient. There were 56 patients tissue and each patient tissue has been sliced into 6-µm thickness so that it has 1, 2, or 3 TMA spot, in total 85 tissue which has been recorded. The TMA were de-waxed and then images were obtained using Spectrum Spotlight 400 FTIR imaging system Spotlight, Perkin Elmer. The images have 6.25-µm spatial resolutions, a 16 cm−1 spectral resolution, and a 15 scan-averaged accumulation in a mid- IR range of 900 to 4000 cm−1, which is spread over the 389 spectral bands.

The FTIR images or Hyperspectral images are mathematically data cubes, which composed of one spectral and two spatial dimensions. Hyperspectral images basically consist of several pixels and each pixel represents the IR spectrum, which ranges from 900 to 4000 cm−1 with 389 wavenumbers (Farah et al., 2016).

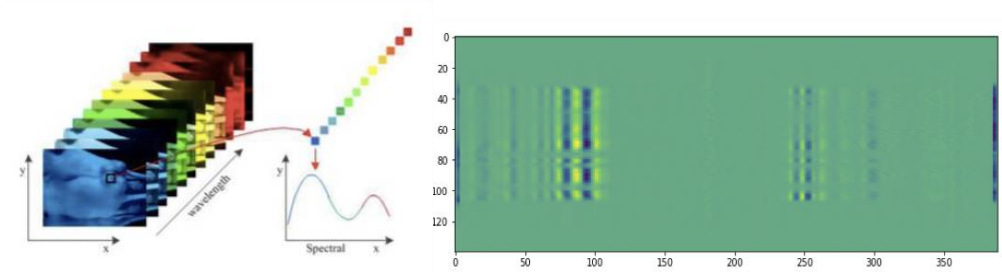
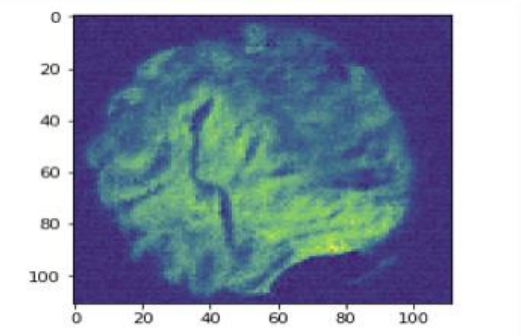


Figure 3.2 Left image shows the hyperspectral image spatial and spectral dimension description with pixel illustration. Retrieved from (Lu, Halig, Wang, Chen, & Fei, 2014). Right image shows the slice of image with one spatial and spectral dimension.



BN	BO	HGD	LGD	N	O	OAC
3	34	3	12	10	16	5

Figure 3.3 Left: Volumetric image of tissue. Right: Class Distribution of Images.

There are 83-labelled images of patient's tissue and there are 7 labels for images those are BN, BO, LGD, HGD, N, O, and OAC. The class distribution of the images is as in Figure 3.3 (Right). Each image consists of [x, y] spatial dimension and 389 spectral bands.

3.2.2 2nd Derivative images

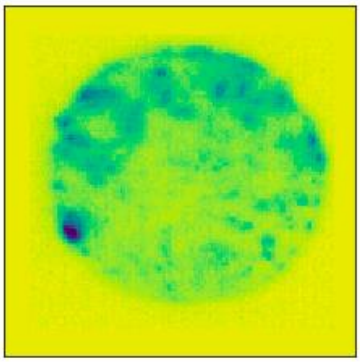


Figure 3.4 2nd Derivative Image

There is another type of dataset also, which was generated by converting raw images into 2nd order Derivative using Satvitzsky-Golay of 9 points, 2nd order polynomial. So now images look like as in figure 3.4, so now this project will deal with two types of images that is

1) Raw Images
2) 2nd Derivative images

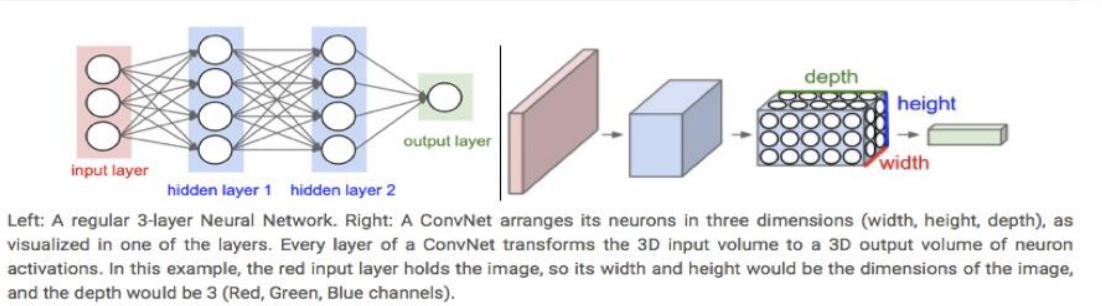
3.3 Data Processing

After getting the dataset, the next step comes with processing of data, which includes loading mat files and removing unnecessary variables from the data, also the size or dimension of images are different so image resizing has been done, and normalization of image pixels has been done in order to eliminate spectral non-uniformity and bring consistency for the training of data. All this has been explained in detail in Chapter 4.

3.4 Model Architecture

This section describes the CNN model architecture, and how each layer of it works in order to detect images. The current work is done by using the basic architecture of the CNN. Convolutional neural network (CNN) architecture is similar to the artificial

neural network as it also narrows downs the output; the difference lies in the arrangement of neurons. In CNN neurons are arranged in 3 Dimension that is **height, width and depth**. Depth basically represents the channels of the spectrum like 3 for RGB and in the data, it's 389.



Left: A regular 3-layer Neural Network. Right: A ConvNet arranges its neurons in three dimensions (width, height, depth), as visualized in one of the layers. Every layer of a ConvNet transforms the 3D input volume to a 3D output volume of neuron activations. In this example, the red input layer holds the image, so its width and height would be the dimensions of the image, and the depth would be 3 (Red, Green, Blue channels).

Figure 3.5 How CNN is different from Artificial Neural Network, image retrieved from http://cs231n.github.io/convolutional-networks/

CNN Architecture mainly consists of three layers: **Convolutional Layer, Pooling Layer and Fully Connected Layer.** Like the artificial neural network, CNN also has Hidden layers where features are detected. It also has a classification part, which classifies and assigns probabilities to the classes. The basic flow of CNN can be explained as follows, there are 3D input images of size n x m are fed to the convolutional layer which uses filters and picks the particular size of image in order to get feature map, its output then fed to the pooling layer which reduces the spatial size of image, output of the pooling is then fed to the fully connected layer where classification of images is been done as final output. Details on each layer and working of CNN has been given below.

3.4.1 Weight Initialization

The first step in CNN is to initialize weights which can be done in several manners like initializing zero weights, random weights or initialize by using Xavier initializer in tensorflow, which automatically adjust the weights to keep the signal in range throughout the network. Biases are also needed to be added like an artificial neural network, which is usually zero.

3.4.2 Convolutional Layer

Convolution, in general, represents a combination of two functions to generate the third function. This concept of convolution is used here, which is done by **Filters**, where the filter is slide over the input image, which is a matrix of pixels. Weights are

multiplied with the matrix in each filter and then all are summed to produce **feature map.**

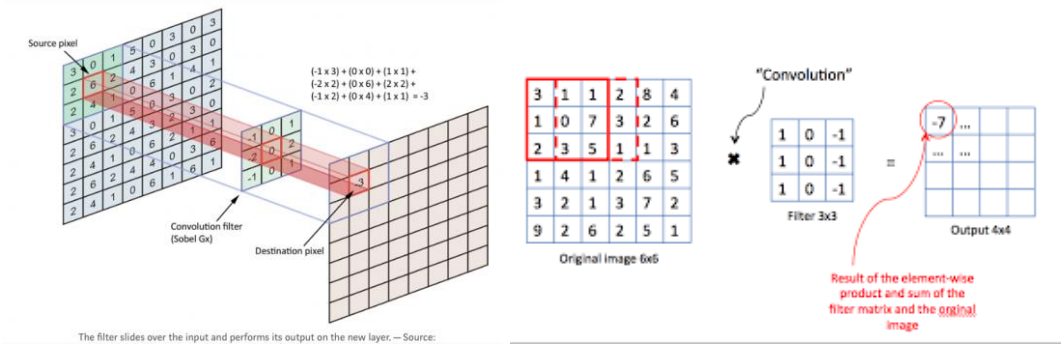


Figure 3.6 Slide of filter window over image, Image retrieved from https://towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2

Stride is the step of each convolutional filter move, like in the previous image it was one when it selected another 3X3 filter (shown with dotted lines). To avoid feature map shrinking, use of padding is done to surround the image with extra zero pixel layer.

Instead of Sigmoid or tanh activation function, ReLU activation function is used to make output non linear, the reason to use ReLU is it does not have gradient vanishing problem unlike sigmoid and tanh, which happens when higher layer units are saturated nearly to -1 or 1 and leading to lower layer to have zero gradients, in result it slows down the optimization convergence process and gives poor local minimum (Maas et. al., 2013):

$$g(z) = max(0, z)$$

It takes 0 for all values below 0 and z for above 0.

3.4.3 Pooling Layer

Pooling layer is used to reduce the spatial size to reduce the dimensionality and computation of feature map and prevents overfitting. It downsamples the volume spatially while keeping the same depth. Each pooling layer corresponds to the previous convolutional layer. The most common pooling operation is max pooling, which is used throughout this thesis. In the below figure, it can be seen that it takes the maximum value from the 2x2 window and downsamples the whole image.

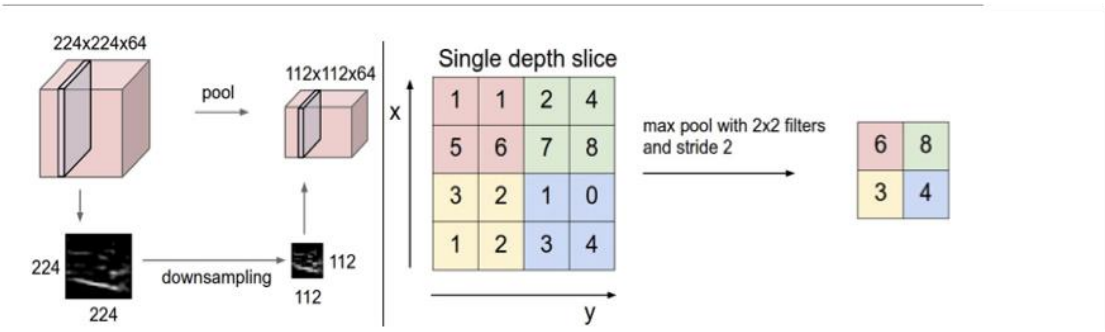


Figure 3.7 Downsampling using Maxpool layer. Image retrieved from http://cs231n.github.io/convolutional-networks/

3.4.4 Fully Connected Layer

Fully Connected Layer (FC) takes the 1D input so 3D matrix needs to be flattened to 1D vector. This is like regular neural network and connected to all neurons of the previous layer. The problem of overfitting can be caused due to a small number of training samples, so some regularization techniques, like L2 regularization and dropout, can be used with the fully connected layer. Dropout is one of the new regularization techniques where it ignores some of the neurons randomly while training the model which reduces the loss function and ultimately reduces the overfitting of the model.

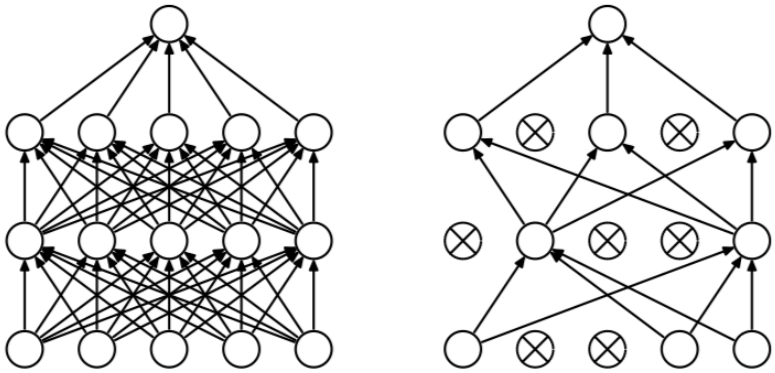


Figure 1: Dropout Neural Net Model. **Left**: A standard neural net with 2 hidden layers. **Right**: An example of a thinned net produced by applying dropout to the network on the left. Crossed units have been dropped.

Figure 3.8 Dropout concept explained by Srivastava et al,2014

After fully connected layer, Softmax along with cross-entropy has been used which maps logits to between 0 and 1, which eventually predict the output, and forces all values to be positive. For the optimization of model Adam optimizer has been used,

unlike stochastic gradient descent, it uses adaptive learning rate for weights in a network with training.

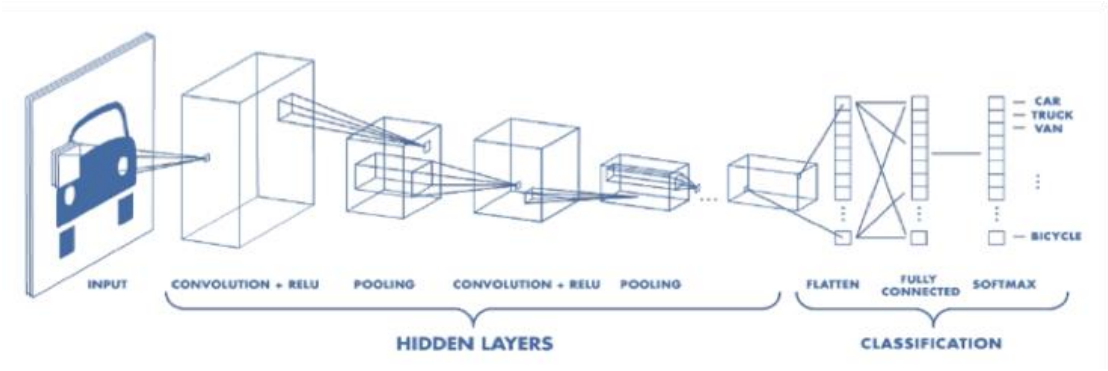


Figure 3.9 Full general CNN Architecture. Image retrieved from https://www.mathworks.com/videos/ introduction-to-deep-learning-what-are-convolutional-neural-networks--1489512765771.html

3.5 Performance Evaluation

For the evaluation of any model, there are several scoring criteria like Accuracy, Precision, Recall and F1 score. To get all this confusion matrix is used which tells how many instances are correctly classified and how many misclassified. On the basis of this count of correctly classified and misclassified, all the scores being calculated. To evaluate this thesis there need to be some criteria to measure the performance of the model. Performance evaluation is highly dependent on the nature of the data, model architecture. For this thesis, main focus remained at F1 score, as due to a high imbalance in data, accuracy can be misleading, as it is highly dependent on the correct prediction that ignores the ratio of classes. In below subsection, all type of metrics has been explained.

3.5.1 Accuracy

It is the ratio of the correct predictions to the total prediction in the test data:

$$Accuracy = \frac{Correct\ Prediction}{Total\ Prediction}$$

It's a good measure if our data is symmetric but as classes in the dataset are highly imbalanced so the use of this as the final performance measure can be misleading though it can be used to see the % of correct predictions.

3.5.2 Precision or Specificity

Precision is the ratio of correct positive prediction or correct prediction of class to the total positive prediction or total prediction of a class.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

3.5.3 Recall or Sensitivity

Recall or sensitivity is the correctly predicted class events.

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

3.5.4 F1 Score

F1 Score is the weighted average of both precision and recall, for uneven classes it works really well as it takes both false positive and false negative in the account.

$$F1\ Score = 2 * \frac{(Recall * Precision)}{(Recall + Precision)}$$

3.6 Summary

This chapter of design and methodology gives the plan and a brief outline of the experiment, which starts with the hypothesis, which actually explains the purpose and focus of this research. Next section provides a detailed overview of the HSI dataset, which will be used for this experiment throughout, it also provides the bits of image generation. Later stages tell the data processing techniques, which will be used in order to feed them to CNN architecture. The Model architecture part after data processing explained the detailed architecture of CNN with working of each layer and benefit of each layer. After that last section gives details about the evaluation methods to be used, on the basis of which the final outcome of the thesis will be decided.

The next section will give a detailed implementation process and results.

CHAPTER 4

4. Implementation and Results

The main aim of this chapter is to give the complete implementation details of Model architecture that was described in the last section and is the basis of this thesis. It will explain different sampling techniques used and what were the results with the use of them. This Chapter will cover the following topics

- Data Preprocessing
- Data Sampling
- Model Architecture
- Results

4.1 Data Preprocessing

It is very important to preprocess the data in order to use them as input in the CNN model, which includes resizing of images, normalization and hot encoding. In our research, we also included two types of images that are raw images and 2nd derivative images, which are mentioned in the previous section. Apart from this, in Figure 3.2 (Right) image, it can be seen that class distribution is highly imbalanced and some classes have only three instances so in order to see the model performances we mostly dealt with two different class distribution that is 7 classes and 3 classes. Three classes will only deal with Normal (N), Oesophageal adenocarcinoma (OAC), and all others are considered as Intermediate stages that are Premature cells (PM). The details on various preprocessing techniques are given below

4.1.1 Resizing of image

All the images have different spatial dimensions due to a different sample size of tissues which varies from 34 -140 in both x and y directions.

To feed images as input to CNN, images need to be of the same dimension as filters are used to extract features from the images, which slides over the dimension of images. Due to this reason, the first step in data pre-processing is to make all images of the same dimension, to do so padding and cropping are two methods.

I. Cropping cannot be done as it can result in loss of information.

II. Padding can be done, where all images are resized to 140 X 140 by taking the maximum dimension image from all the images and adding zero pixels along the matrix of pixels to the left, right, top and bottom of the image and keeping the colour channels of the same size that is 389.

The padding has one drawback in that it can add noise to the images, which can impact the performance. Though choosing the appropriate size for padding can be used and the noise effect can be ignored.

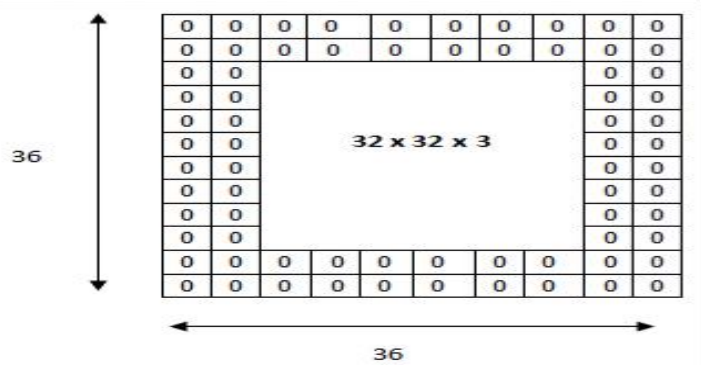


Figure 4.1 Example of zero padding, where 32x32x3 converted into 36x36x3. Image retrieved from https://www.kaggle.com/pouryaayria/convolutional-neural-networks-tutorial-tensorflow

Minimum and maximum dimension images are: (36 X 42) and (140 X 136).

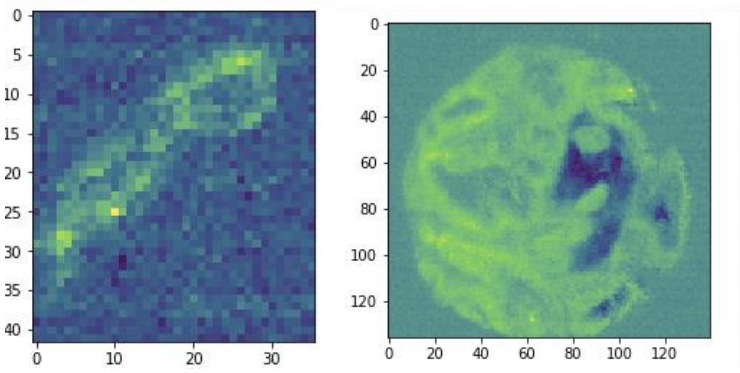


Figure 4.2 Left: Image shows the minimum dimension image that is 36x42. Right: shows the maximum size image that is 140x136

Dimension of all images are then converted to 140 X 140 as described in Figure 4.2

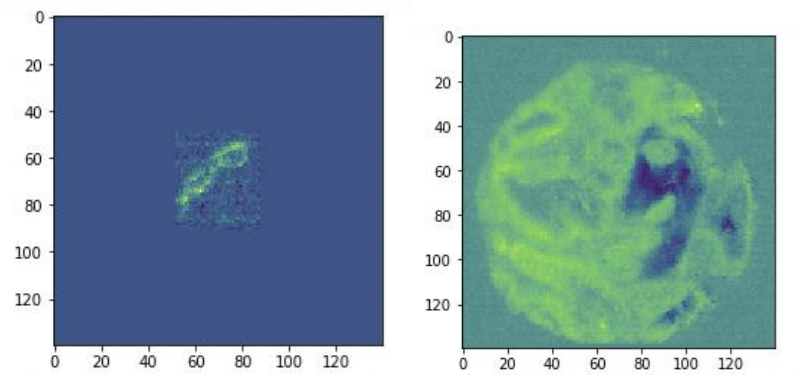


Figure 4.3 Left: 36x42 image converted to 140x140. Right: 140x136 converted to 140x140

4.1.2 Normalization

Data normalization of each patient image matrix has been done to eliminate spectral non-uniformity and bring consistency in the dynamic range of pixels. The 3-D image of the range of $I: \{X \in R\} \rightarrow \{Min \dots Max\}$ is converted to range of $Inew: \{X \in R\} \rightarrow \{newMin \dots newMax\}$ by using linear normalization.

$$Inew = (I - Min)\frac{newMax - newMin}{Max - Min} + newMin$$

Using the normalization technique, all patient image matrices are converted between -1 and 1.

4.1.3 Hot Encoding

As the data is categorical data, hot encoding or conversion into numerical is required. But before hot encoding label encoding need to be done as data categories are in string format. Label encoding converts the unique categories into the integer range. Like here, all 7 categories convert into 0-6 ranges, which can be seen as below.

['BN'-0, 'BO'-1, 'HGD'-2, 'LGD'-3, 'N'-4, 'O'-5, 'OAC'-6]

After label encoding one hot encoding is done, which converts the labels into binary values like below in Table 4.1.

BN	BO	HGD	LGD	N	O	OAC
1.	0.	0.	0.	0.	0.	0.
0.	1.	0.	0.	0.	0.	0.
0.	0.	1.	0.	0.	0.	0.
0.	0.	0.	1.	0.	0.	0.
0.	0.	0.	0.	1.	0.	0.
0.	0.	0.	0.	0.	1.	0.
0.	0.	0.	0.	0.	0.	1.

Table 4.1 shows the label encoding of categories

Similarly for three class data we will have label encoding as N-0, OAC-1 and PM-2.

4.2 Data Sampling

In the experiment, a stratified random sampling technique has been used which splits data into train, test and validation data class wise, which make sure that all classes are there in the three divisions. Basic sampling strategy has been used as 20% of Test dataset whereas Validation dataset has been taken in a manner that one instance from each class has been taken; remaining data has been taken for training data. So data distribution of three parts looks as follows for raw data.

Training	59
Test	17
Validation	7
Total	83

Table 4.2 Data distribution among Train, Test and Validation

In later implementations the count of instances varies as per the technique used.

4.3 Model Architecture Implementation

The basic architecture of CNN has been used for the implementation which consists of one convolutional layer, max-pooling layer and one fully connected layer. In this

thesis, along with basic architecture ReLU activation function, Dropout, and Softmax have been used.

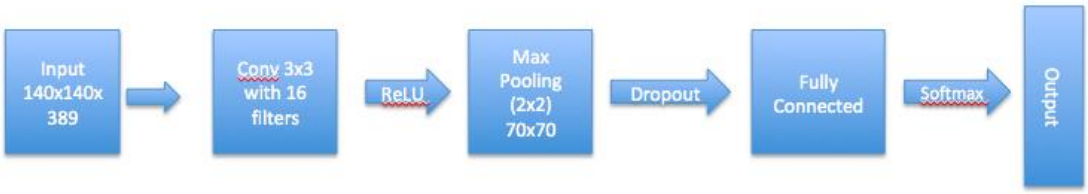


Figure 4.4 Experiment Model Architecture

4.3.1 Model A- Baseline model with 7 classes raw images

Architecture chose is the basic model that is in figure 4.5:

INPUT -> [CONV -> RELU -> POOL] -> DROPOUT -> [FC [SEP]->SOFTMAX]

In the model at Convolutional Layer 16 filters of the 3x3 window has been used. For the Adam optimizer, the exponential decay of learning rate has been used with an initial rate of 0.05. As training data size is too small, the batch size of 8 has been used along with dropout of 0.25.

As the input size is 140x140, max pooling downsamples the image pixel to 70x70, which in return provides 78400 features. The features are then fed to the fully connected layer with the dropout of 0.25, the fully connected output then uses maps logits to between 0 and 1, which gives the final output of 7 classes. The cost or error loss has been calculated using cross entropy. The model has been trained with total 250-iterations.

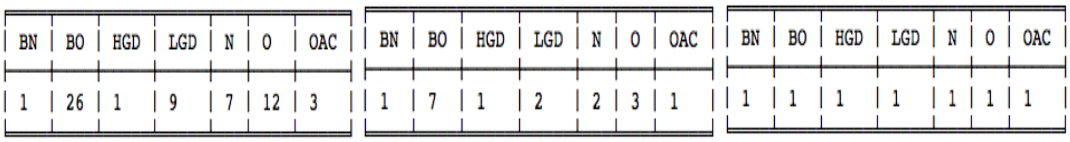
For this model, raw images with 7 classes have been used and sampling done for it is the same as in table 4.2. This will be the base model, as it will be used to see the improvement of the model on changing the sampling technique and hyperparameters.

BN	BO	HGD	LGD	N	O	OAC
3	34	3	12	10	16	5

Figure 4.5 Total raw image class distribution

Seven classes in the data are as follows BN, BO, HGD, LGD, N, O, and OAC. The main issue with this data is some classes have very few instances to divide into three parts, due to this for few classes only one instance remains in training data.

The above distribution clearly shows that BN and HGD has only three instances, while OAC that is cancer class has only 5 instances.



BN	BO	HGD	LGD	N	O	OAC
1	26	1	9	7	12	3

BN	BO	HGD	LGD	N	O	OAC
1	7	1	2	2	3	1

BN	BO	HGD	LGD	N	O	OAC
1	1	1	1	1	1	1

Figure 4.6 Distribution of Train, Test, and Validation data from left to right

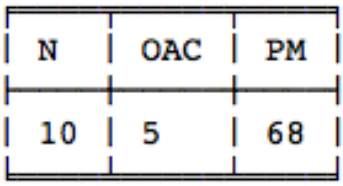
The above distribution shows the instances of each class in training, test and validation data.

4.3.2 Model B- Model with 7 classes and 2nd derivative images

For this model, the architecture, hyperparameters and sampling technique remained same as Model A. The only change is the image type that is now the same model is fed with the 2nd derivative images instead of raw images.

4.3.3 Model C- Model with 3 classes for both raw images and 2nd derivative images

For this model, the class number has been reduced to three by combining all five intermediate stages that is BN, BO, HGD, LGD, O into one class that is **Premature (PM)** and keeping Normal (N) and OAC as it is.



N	OAC	PM
10	5	68

Figure 4.7 Count of instances of each class

The class distribution in data looks like as in Figure 4.6, which clearly shows that there is an imbalance class distribution that is PM has 68 instances, N has only 10 instances while OAC has 5 instances. The sampling was done in a similar manner as in the other two model.

The sampling for the data has been done in the same fashion as in the previous model, just due to 3 classes now training data has 63 instances while test data is 17 and validation data has 3 instances.

4.3.4 Model D- Model with Synthetic Minority Oversampling Technique (SMOTE)

This model uses Synthetic Minority Oversampling (SMOTE). In SMOTE, it identifies the minority classes and creates the synthetic samples of the instances in equal to majority class rather than simply duplicating or resampling with replacement.

SMOTE has been applied to only training data so that test data remains untouched. Now training data changes from 63 instances to 159 as 96 synthetic examples of N and OAC has been added to the data.

For this, both raw images and 2nd derivative images with 3 classes have been used to see if it can identify instance correctly. Apart from that, all parameters kept same as in the baseline model.

4.3.5 Model E- Model with image augmentation

4.3.5.1 Rotating 45 Degree

For this model, the four images of minority classes have been rotated by 45-degree difference till 315 degree that is now we have each image of N and OAC with 45, 90, 135, 180, 225, 270, and 315 degrees, which increased the minority image data for training data. So finally the new rotated images are 28 new clones of Normal class images and 21 clones of OAC class images. Kept test and validation data same as previous models that are 17 test instances and 3 validation instances. With this, the class distribution of training data will be as follows:

N	OAC	PM	Total
35	24	53	112

Table 4.3 Class wise training data instance count

4.3.5.2 Rotating 30 degree

The four images of minority classes have also been rotated by 30-degree difference till 315 degree that is now we have each image of N and OAC with 30,60, 90, 120, 160, 190, 220, 250, 280, 310 and 340 degree, which increased the minority image data for training data. So finally the new rotated images are 44 new clones of Normal class images and 33 clones of OAC class images. Kept test and validation data same as previous models that are 17 test instances and 3 validation instances. With this, the class distribution of training data will be as follows:

N	OAC	PM	Total
51	36	53	140

Table 4.4 Class wise training data instance count

4.4 Results

This thesis revolves around the experiments with a try to improve the prediction by manipulating data and changing hyperparameters to see the changes in the model. Throughout the evaluation, the confusion matrix and f1 score have been the criteria to know how the model performed. Confusion Matrix shows the count of correct prediction along with a count of incorrect predictions.

4.4.1 Model A- Baseline model with 7 classes raw images

The baseline model result is the base result by which all other models are being compared.

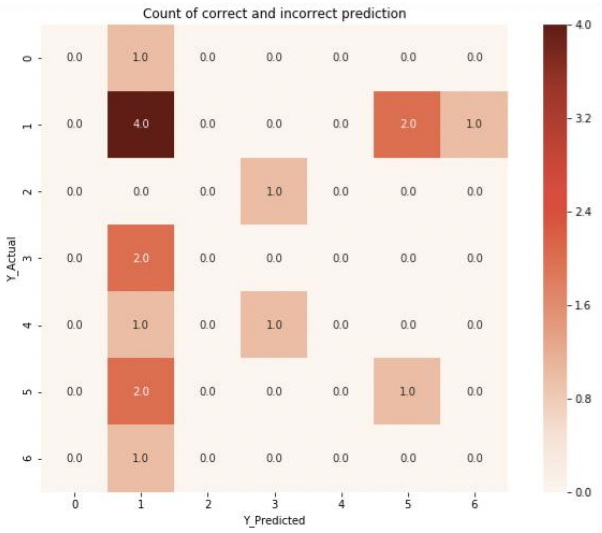


Figure 4.8 CM for Model A

The confusion matrix (CM) of the model shows the count of predictions among 7 classes. The 0-6 values represent as follows

[BN-0, BO-1, HGD-2, LGD-3, N-4, O-5, OAC-6]

It clearly shows that it correctly predicted 4 BO classified images out of 7 and 1 O classified image out of 3.

So individual accuracy of classes is as follows:

BN	BO	LGD	HGD	N	O	OAC
0%	57.14%	0%	0%	0%	33.33%	0%

Table 4.5 Percentage of correct prediction

Apart from individual accuracy final accuracy came out to be 29.41% whereas Precision is 0.208556149733, Recall 0.294117647059, and **F1 Score** is **0.241830065359**.

4.4.2 Model B- Model with 7 classes and 2nd derivative images

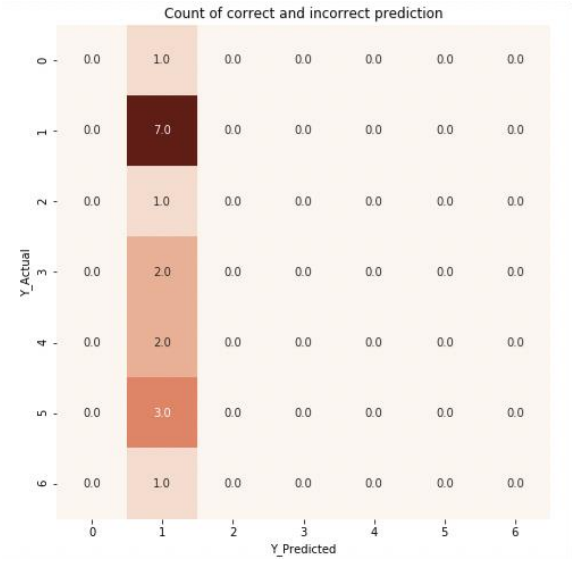


Figure 4.9 CM for Model B

It clearly shows that it is able to identify Only BO correctly, apart from that it misclassified all other classes.

So individual accuracy for BO is 100% while for all other classes it is 0%. The Overall accuracy came out to be 41.18% which is greater than Model I whereas Precision is 0.16955017301 which is lower than the baseline model that is Model A, **F1 Score** is **0.24**, which shows no difference from Model A. This result actually gives the impression that 2nd derivative images don't make any difference in identifying any kind of in case of 7 classes tissue and these results are the sign of how imbalance data can impact the results and makes it difficult to identify any of the minority classes.

4.4.3 Model C- Model with 3 classes for both raw images and 2nd derivative images

The confusion matrix of the model shows the count of predictions among 3 classes. The 0-2 values represents as follows [‘N'–0, ‘OAC'-1,'PM'-2]. Looking at the confusion matrix it shows that PM which has majority instances are predicted correctly apart from that all classes classified incorrectly**.**

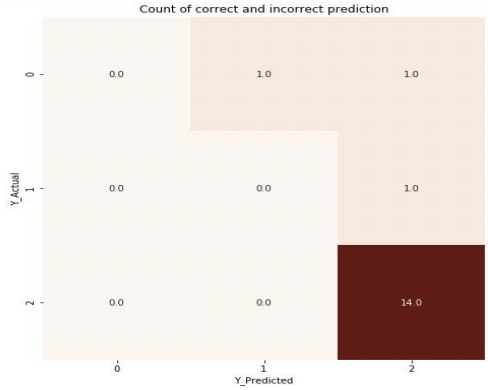


Figure 4.10 CM for model C (raw images)

The overall accuracy came out to be 82.35%, which is greater than other baseline models A, whereas Precision is 0.72, which is greater than model A, F1 Score is 0.768. Though this score is mainly due to the majority class whereas minority class are not able to classify.

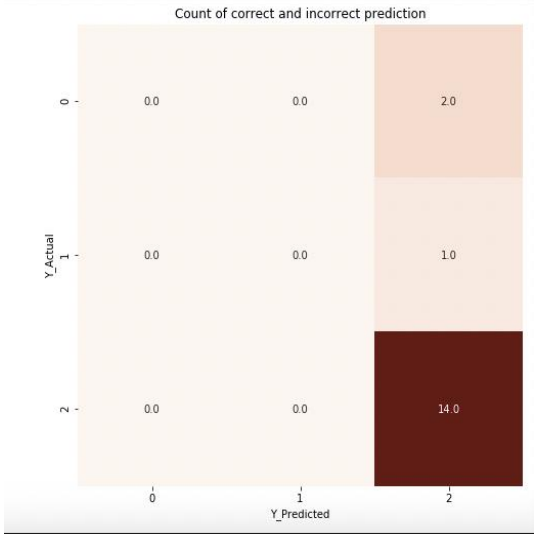


Figure 4.11 CM of Model C with 2nd Derivative image

The confusion matrix for the 2nd derivative shows the similar result as in with raw images. The overall accuracy came out to be 82.35%, whereas Precision is 0.678, which is lower than raw images, while **F1 Score** is **0.744**. This result clearly shows that raw images are similar to 2nd derivative images and clearly this doesn't impact the prediction power of the model.

4.4.4 Model D- Model with Synthetic Minority Oversampling Technique (SMOTE)

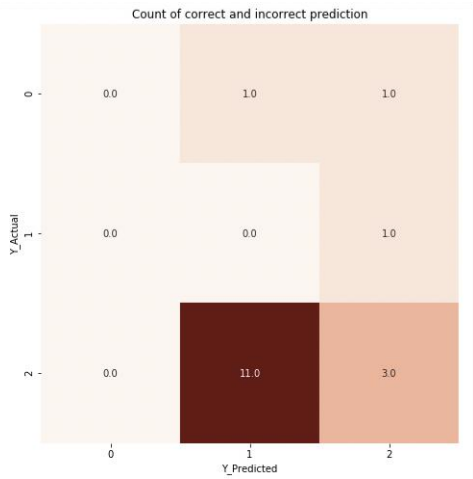


Figure 4.12 CM of Model D

Result clearly shows that it misclassified most of the instances and only 3 instances were correctly classified as PM. The overall accuracy came out to be 17.65% whereas Precision came out to be 0.494 and **F1 Score** is **0.26,** which came out to be lower than other models.

This result came out to be the same for both raw images and 2nd derivative image. But result clearly shows that in this case SMOTE oversampling couldn't help, it increased the data reasonably but still, synthetic samples did not help the classifier to train and classify the classes.

4.4.5 Model E- Model with image augmentation

4.4.5.1 With 45 degree rotation

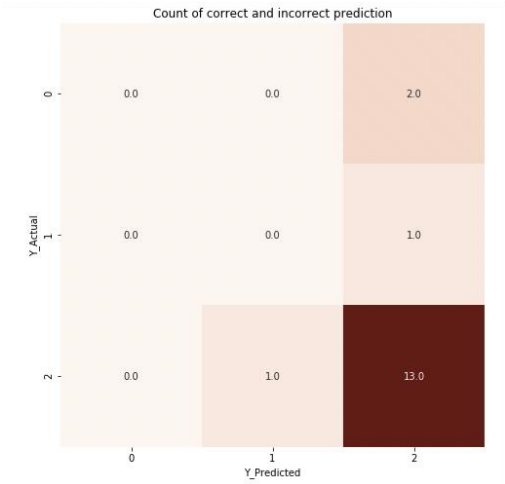


Figure 4.13 CM with image augmented to 45 degree

The Overall accuracy came out to be 76.47%, while Precision is 0.669 and **F1 Score** is **0.71**. The result does not show much variation from previous models.

4.4.5.2 With 30 degree rotation

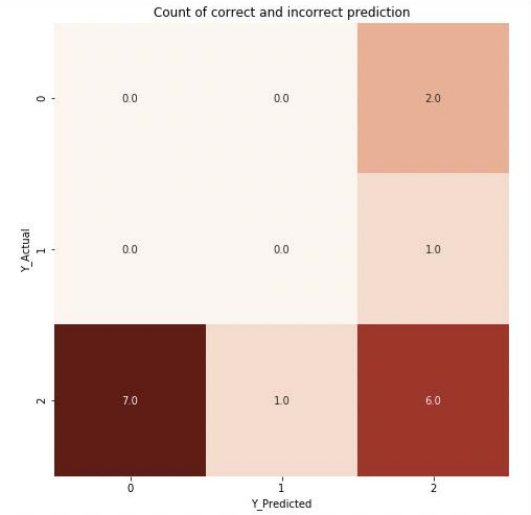


Figure 4.14 CM with image augmented to 30 degree

The Overall accuracy came out to be 41%, while Precision is 0.55 and **F1 Score** is **0.43**, which is very lower than when images were rotated at 45 degree.

The whole result with image augmentation didn't show any good result though it misclassified some of the majority classes with the minority classes. In this, only raw images have been used, as result with 2nd derivative images lower than with raw images.

Model	Model Description	Accuracy	F1Score
A	Baseline Model with raw images, 7 classes, Batch Size 8 and dropout 0.25	29.41%	0.242
B	Model with 2nd Derivative images, 7 classes, Batch Size 8 and dropout 0.25	41.18%	0.24
C	Model with raw images, 3 classes, Batch Size 8 and dropout 0.25	82.35%	0.768 (raw) 0.744 (2nd derivative)
D	Model with SMOTE, raw images, 3 classes, Batch Size 8 and dropout 0.25	17.65%	0.26
E (with 45 degree)	Image augmentation of N and OAC with 45 degree rotation of raw images angles 45,90,135,180,225,270, and 315 - batch size -8 and dropout 0.25	76.47%	0.71
E(with 30 degree)	Image augmentation of N and OAC with 30 degree rotation of raw images angles 30,60,90,120,150,180,210,240,270,300 and 330 - batch size -8 and dropout 0.25	41%	0.43
E (with 45 degree)	Image augmentation of N, OAC, classes of raw images angles 45,90,135,180,225,270, and 315 - batch size -16 and dropout 0.25	64.71%	0.67

Table 4.6 Model results with different techniques

CHAPTER 5

5. Evaluation/Analysis

5.1 Evaluation of Results

F1 Score has been the evaluation criteria throughout the thesis, in the baseline model it was evident that it gave very bad results in term of F1 score and accuracy, but that is purely because of having so many classes and very fewer instances per class to train the CNN model. Once the classes reduced to 3 from 7, the results took a jump from the base model's result, but that could be because of overfitting of majority classes rather than identifying the other two classes that are Normal and OAC classes which were having really fewer instances. But it can be seen that it detected all the premature stages correctly. Use of oversampling technique SMOTE with the same hyperparameters didn't show good results, instead, it really reduced the F1 Score. Another oversampling technique used is Image augmentation by rotating the images by 45 degrees and 30 degrees, for 45-degree rotation result came out to be same as 3 classes model whereas 30-degree rotation didn't show that improvement in predictions.

So far as the underlying research question regarding the predictive power of a CNN model with HSI on fine tuning, the conclusion is that ***valid sampling techniques and hyperparameters, can impact the performance of the model.*** Though the major problem with the results and technique was the data imbalance, so results were not great, but in future, with more hyperparameters and sampling technique better results can be acquired.

5.2 Strengths of the Result

The strength of the results is the ability to know that other fine-tuning techniques can be used in order to improve the results. As throughout the thesis results were impacted due to the change in techniques. It had the impact which can help in the future work of research.

Also, it shows how imbalanced data directly impacts the performance of the classifier. And in the future by having more data we can improve the results.

Also, this thesis opened gate for more techniques, which can be used in terms of model architecture, or feature extraction or changing hyperparameters.

5.3 Limitation of the results

Due to the time constraints and slow training process, it was difficult to train the model with different hyperparameters like the batch size and epochs for the same sampling techniques those were used, which might show the better results. Also for the same reason, other techniques like CNN with PCA, ensembled model, multiple convolutional architectures were not able to use.

The results were not stable throughout the thesis due to the time is taken by the model and obstacle comes between in the long run. Many times the system got crashed due to generating data after each type of sampling.

Results came out to be more biased towards the majority classes and other classes had 0% accuracy for the perfect reason for data imbalance and oversampling didn't help.

CHAPTER 6

6. Conclusion

6.1 Research and Experiment Overview

This research formulated, built and evaluated the basic CNN algorithm for performing the classification of the tissues. This research is the contribution towards the detection of the early stage of oesophageal cancer, to help the patient to get endoscopy surveillance, surgery or radiation treatment requirements before it reaches to the OAC level. This research has been carried out in a way that in the beginning literature review carried out in order to see the implication of deep learning in the field of medical, also how hyperspectral images are being helpful in the field of medical imaging and analysis.

In order to build the model based on CNN algorithm, a thorough experiment was performed. Initially, the basic model has been built, which was then fine-tuned to see the improvement in the results using performance criteria of accuracy and F1 score. Experiments performed by changing the sampling techniques such as SMOTE, Image augmentation or changing the number of classes in order to get better results for the classifier.

The technique, which gave the best result, was Model C in terms of F1 Score that was 0.76. Though in this result it is only identifying the class with majority instances, after balancing the data by oversampling techniques like SMOTE or Image augmentation the image augmentation that is Model E (with 45-degree rotation) performed well with the F1 score of 0.71 that was near to the model C results. For the implementation of our architecture, we used python with Tensorflow, in the whole experiment basic implementation code has been taken from Thomalm work through network architecture and fine-tuning is done by self.

6.2 Future work and recommendations

This project focuses on CNN algorithm to perform the image classification task, from the experiment, it is found that tuning the model such as changing the sampling techniques impacts the classification accuracy and F1 Score. The future work can involve model ensembling architecture, more changes in hyperparameters like learning

rate, batch size and dropout rate and more different sampling techniques can be used in order to improve the result.

Also with the use of GPU, the model can be trained with more epochs, that may improve the result.

The dataset used in this project is very limited and it is trained using 83 images only due to availability and confidentiality of data. Also, the biggest constraint for this project was the data imbalance, which builds the biased classifier rather than a robust classifier. Getting more images of the minority class and use those images to build the model can do further work. Also the samples of images we got had different dimensions and in some images, dimension difference was too much which while resizing the images added noise to the smaller dimension images, So if we can get data with bit of similar size it can reduce the problem of noise in the data and could have significant impact on the performance of the model. Also the CNN architecture implementation, we didn't apply any feature extraction techniques, the future work can also involve changing this by adding PCA or some other feature extraction techniques to compare the result and see if they provide the better performance.

Also use of proper error analysis to optimize the model like plotting the error rate with epoch or validation error with training error can provide much more insight into the working of the model. These techniques could help in analyzing the model in a clear manner and could help in avoiding the biases of the model.

7. References

Andreasen, N. C., Harris, G., Cizadlo, T., Arndt, S., O'Leary, D. S., Swayze, V., & Flaum, M. (1994). Techniques for measuring sulcal/gyral patterns in the brain as visualized through magnetic resonance scanning: BRAINPLOT and BRAINMAP. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(1), 93–97.

Akbari, H., Halig, L. V., Schuster, D. M., Osunkoya, A., Master, V., Nieh, P. T., … Fei, B. (2012). Hyperspectral imaging and quantitative analysis for prostate cancer detection. *Journal of Biomedical Optics*, *17*(7), 0760051. https://doi.org/10.1117/1.JBO.17.7.076005

Besharat, S., Jabbari, A., Semnani, S., Keshtkar, A., & Marjani, J. (2008). Inoperable esophageal cancer and outcome of palliative care. *World Journal of Gastroenterology*, *14*(23), 3725. https://doi.org/10.3748/wjg.14.3725

Brown, M. P. S., Grundy, W. N., Lin, D., Cristianini, N., Sugnet, C. W., Furey, T. S., … Haussler, D. (2000). Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proceedings of the National Academy of Sciences*, *97*(1), 262–267. https://doi.org/10.1073/pnas.97.1.262

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Corley, D. A., Levin, T. R., Habel, L. A., Weiss, N. S., & Buffler, P. A. (2002). Surveillance and survival in Barrett's adenocarcinomas: A population-based study. *Gastroenterology*, *122*(3), 633–640. https://doi.org/10.1053/gast.2002.31879

Cruz-Roa, A. A., Arevalo Ovalle, J. E., Madabhushi, A., & González Osorio, F. A. (2013). A Deep Learning Architecture for Image Representation, Visual

Interpretability and Automated Basal-Cell Carcinoma Cancer Detection. In K. Mori, I. Sakuma, Y. Sato, C. Barillot, & N. Navab (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013* (Vol. 8150, pp. 403–410). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-40763-5_50

Danaee, P., Ghaeini, R., & Hendrix, D. A. (2017). A DEEP LEARNING APPROACH FOR CANCER DETECTION AND RELEVANT GENE IDENTIFICATION (pp. 219–229). WORLD SCIENTIFIC. https://doi.org/10.1142/9789813207813_0022

Duncan, J. S., & Ayache, N. (2000). Medical image analysis: progress over two decades and the challenges ahead. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(1), 85–106. https://doi.org/10.1109/34.824822

Farah, I., Nguyen, T. N. Q., Groh, A., Guenot, D., Jeannesson, P., & Gobinet, C. (2016). Development of a memetic clustering algorithm for optimal spectral histology: application to FTIR images of normal human colon. *The Analyst*, *141*(11), 3296–3304. https://doi.org/10.1039/C5AN02227D

Feber, A., Xi, L., Luketich, J. D., Pennathur, A., Landreneau, R. J., Wu, M., … Litle, V. R. (2008). MicroRNA expression profiles of esophageal cancer. *The Journal of Thoracic and Cardiovascular Surgery*, *135*(2), 255–260. https://doi.org/10.1016/j.jtcvs.2007.08.055

Ferris, D. G., Lawhead, R. A., Dickman, E. D., Holtzapple, N., Miller, J. A., Grogan, S., … Faupel, M. L. (2001). Multimodal Hyperspectral Imaging for the Noninvasive Diagnosis of Cervical Neoplasia. *Journal of Lower Genital Tract Disease*, *5*(2), 65–72. https://doi.org/10.1046/j.1526-0976.2001.005002065.x

Gurcan, M. N., Boucheron, L. E., Can, A., Madabhushi, A., Rajpoot, N. M., & Yener, B. (2009). Histopathological Image Analysis: A Review. *IEEE Reviews in Biomedical Engineering*, *2*, 147–171. https://doi.org/10.1109/RBME.2009.2034865

Halicek, M., Lu, G., Little, J. V., Wang, X., Patel, M., Griffith, C. C., … Fei, B. (2017). Deep convolutional neural networks for classifying head and neck cancer using hyperspectral imaging. *Journal of Biomedical Optics*, *22*(6), 060503. https://doi.org/10.1117/1.JBO.22.6.060503

Hu, W., Huang, Y., Wei, L., Zhang, F., & Li, H. (2015). Deep Convolutional Neural Networks for Hyperspectral Image Classification. *Journal of Sensors*, *2015*, 1–12. https://doi.org/10.1155/2015/258619

J. Sklansky and D. Ballard, aTumor Detection in Radiographs,o Computers and Biomedical Research, vol. 6, no. 4, pp. 299-321, Aug. 1973. https://doi.org/10.1016/0010-4809(73)90066-9

Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. *ArXiv:1412.6980 [Cs]*. Retrieved from http://arxiv.org/abs/1412.6980

Li, Q., He, X., Wang, Y., Liu, H., Xu, D., & Guo, F. (2013). Review of spectral imaging technology in biomedical engineering: achievements and challenges. *Journal of Biomedical Optics*, *18*(10), 100901. https://doi.org/10.1117/1.JBO.18.10.100901

Lu, G., Qin, X., Wang, D., Chen, Z. G., & Fei, B. (2015). Quantitative wavelength analysis and image classification for intraoperative cancer diagnosis with hyperspectral imaging. In R. J. Webster & Z. R. Yaniv (Eds.) (p. 94151B). https://doi.org/10.1117/12.2082284

Lu, G., & Fei, B. (2014). Medical hyperspectral imaging: a review. *Journal of Biomedical Optics*, *19*(1), 010901. https://doi.org/10.1117/1.JBO.19.1.010901

Maas, A. L., Hannun, A. Y., & Ng, A. Y. (2013, June). Rectifier nonlinearities improve neural network acoustic models. In *Proc. icml* (Vol. 30, No. 1, p. 3).

Martin, M. E., Wabuyele, M. B., Chen, K., Kasili, P., Panjehpour, M., Phan, M., … Vo-Dinh, T. (2006). Development of an Advanced Hyperspectral Imaging (HSI) System with Applications for Cancer Detection. *Annals of Biomedical Engineering*, *34*(6), 1061–1068. https://doi.org/10.1007/s10439-006-9121-9

Makantasis, K., Karantzalos, K., Doulamis, A., & Doulamis, N. (2015). Deep supervised learning for hyperspectral data classification through convolutional neural networks (pp. 4959–4962). IEEE. https://doi.org/10.1109/IGARSS.2015.7326945

Mei, S., Ji, J., Hou, J., Li, X., & Du, Q. (2017). Learning Sensor-Specific Spatial-Spectral Features of Hyperspectral Images via Convolutional Neural Networks. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(8), 4520–4533. https://doi.org/10.1109/TGRS.2017.2693346

Musto, P., Calarco, A., Pannico, M., La Manna, P., Margarucci, S., Tafuri, A., & Peluso, G. (2017). Hyperspectral Raman imaging of human prostatic cells: An attempt to differentiate normal and malignant cell lines by univariate and multivariate data analysis. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *173*, 476–488. https://doi.org/10.1016/j.saa.2016.09.034

Sahiner, B., Heang-Ping Chan, Petrick, N., Datong Wei, Helvie, M. A., Adler, D. D., & Goodsitt, M. M. (1996). Classification of mass and normal breast tissue: a convolution neural network classifier with spatial domain and texture images. *IEEE Transactions on Medical Imaging*, *15*(5), 598–610. https://doi.org/10.1109/42.538937

Schultz, R. A., Nielsen, T., Zavaleta, J. R., Ruch, R., Wyatt, R., & Garner, H. R. (2001). Hyperspectral imaging: A novel approach for microscopic analysis. *Cytometry*, *43*(4), 239–247. https://doi.org/10.1002/1097-0320(20010401)43:4<239::AID-CYTO1056>3.0.CO;2-Z

Sharma, H., Zerbe, N., Klempert, I., Hellwich, O., & Hufnagl, P. (2017). Deep convolutional neural networks for automatic classification of gastric carcinoma using whole slide images in digital histopathology. *Computerized Medical Imaging and Graphics*, *61*, 2–13. https://doi.org/10.1016/j.compmedimag.2017.06.001

Shiozaki, H., Tahara, H., Kobayashi, K., Yano, H., Tamura, S., Imamoto, H., … Mori, T. (1990). Endoscopic screening of early esophageal cancer with the Lugol dye method in patients with head and neck cancers. *Cancer*, *66*(10), 2068–2071. https://doi.org/10.1002/1097-0142(19901115)66:10<2068::AID-CNCR2820661005>3.0.CO;2-W

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, *15*(1), 1929-1958.

Siddiqi, A. M., Li, H., Faruque, F., Williams, W., Lai, K., Hughson, M., … Johnson, W. (2008). Use of hyperspectral imaging to distinguish normal, precancerous, and cancerous cells. *Cancer*, *114*(1), 13–21. https://doi.org/10.1002/cncr.23286

Wang, H., Cruz-Roa, A., Basavanhally, A., Gilmore, H., Shih, N., Feldman, M., … Madabhushi, A. (2014). Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features. *Journal of Medical Imaging*, *1*(3), 034003. https://doi.org/10.1117/1.JMI.1.3.034003

Wang, Q., Wang, J., Zhou, M., Li, Q., & Wang, Y. (2017). Spectral-spatial feature-based neural network method for acute lymphoblastic leukemia cell identification via microscopic hyperspectral imaging technology. *Biomedical Optics Express*, *8*(6), 3017. https://doi.org/10.1364/BOE.8.003017

Yang, X., Yeo, S.-Y., Hong, J. M., Wong, S. T., Tang, W. T., Wu, Z. Z., … Su, Y. (2016). A Deep Learning Approach for Tumor Tissue Image Classification. ACTAPRESS. https://doi.org/10.2316/P.2016.832-025

"CS231n Convolutional Neural Networks for Visual Recognition." [Online]. Available: http://cs231n.github.io/transfer-learning/. [Accessed: 29- May-2018]