

A DEEP LEARNING BASED PIPELINE FOR METASTATIC BREAST  
CANCER CLASSIFICATION FROM WHOLE SLIDE IMAGES (WSI)

by

ARJUN PUNABHAI VEKARIYA

Presented to the Faculty of the Graduate School of  
The University of Texas at Arlington in Partial Fulfillment  
of the Requirements  
for the Degree of

MASTER OF SCIENCE IN COMPUTER SCIENCE AND ENGINEERING

THE UNIVERSITY OF TEXAS AT ARLINGTON  
May 2017

Copyright © by ARJUN PUNABHAI VEKARIYA 2017

All Rights Reserved

To my parents and my siblings.

## ACKNOWLEDGEMENTS

I wish to express my sincere thanks to my supervising professor, Dr. Junzhou Huang who encouraged me to do this thesis without whom this thesis would not have been possible. His irreversible fortitude and constant motivation are the main reasons for the successful outcomes of my research. I sincerely express my gratitude to Dr. Heng Huang and Dr. Jia Rao for spending their valuable time by serving on my committee.

I would like to thank Mr. Sheng Wang, my P.hd mentor for his continuous assistance in this thesis and also I wish to mention special thanks to Mr. Zhifei Deng, Mr. Zheng Xu, Mr. Jiawen Yao, Mr. Ashwin Raju and other friends in my lab for constantly motivating me to achieve success and sharing their suggestions on every step in this project.

I would fail if I forget to express my gratitude to my friends Mr. Neeraj Mishra, Mr. Aman Nagar, Mr. Vaibhav Sharma, Mr. Mayur Murlidhar, Mr. Shreyash, Mr. Sanjay, Mr. Karan and all my dear friends at Arlington and India for their boundless belief in my abilities and endless encouragement for my success.

Above all I express my earnest thanks to my dear parents, my brother, my sisters, my cousins, far and near family for all their blessings and sacrifices they had done to help chase my dreams and live my passion. Finally, I thank God for all the opportunities he creates for me.

April 19, 2017

## ABSTRACT

# A DEEP LEARNING BASED PIPELINE FOR METASTATIC BREAST CANCER CLASSIFICATION FROM WHOLE SLIDE IMAGES (WSI)

ARJUN PUNABHAI VEKARIYA, M.S.

The University of Texas at Arlington, 2017

Supervising Professor: Dr. Junzhou Huang

Cancer, the second deadliest diseases on the planet, is a generalized term for the class of diseases caused by the proliferation of abnormal cells in a human body. These abnormal cells are caused due to unwanted growth of new cells and improper recycling of old or damaged cells. Diagnosis methods available today in medical industry are very time-consuming as pathologist has to manually analyze sentinel lymph nodes, which requires scanning entire whole slide image for detecting metastasis region. Efforts have made in developing faster computer-aided methods for analyzing whole slide images but historical approaches have focused primarily on low-level image analysis tasks (e.g., color normalization, nuclear segmentation, and feature extraction) so they are not generalized, thus not useful for practical use in clinical practices.

In this thesis, a Deep Learning based classification pipeline for detection of cancer metastases from histological images is proposed. The pipeline consists of five stages: 1. Region of Interest (ROI) detection with Image processing. 2. Tiling ROI. 3. Deep Convolutional Neural Network (ConvNet) for tile-based classification. 4. Building tumor probability heat-maps. 5. Post-processing of heat-maps for slide-

based classification. GoogLeNet, a deep 27 layer ConvNet is used to distinguish positive tumor areas from negative ones. The key challenge of identifying hard negative areas (areas surrounding tumor region) is tackled with ensemble learning approach using two Deep ConvNet models. Using dataset of the Camelyon'16 grand challenge, the proposed pipeline achieves an area under the receiver operating curve (ROC) of 92.57%, which beats the winning method of Camelyon'16 grand challenge, developed together by Harvard & MIT research laboratories. This results reflect the potential of using deep learning to produce significant improvements in the accuracy of pathological diagnoses.

## TABLE OF CONTENTS

ACKNOWLEDGEMENTS . . . . .	iv
ABSTRACT . . . . .	v
LIST OF ILLUSTRATIONS . . . . .	ix
LIST OF TABLES . . . . .	xii
Chapter	Page
1. INTRODUCTION . . . . .	1
1.1 Cancer . . . . .	1
1.2 Breast cancer . . . . .	3
1.2.1 Types of Breast cancer . . . . .	3
1.2.2 Diagnosis . . . . .	5
1.2.3 Treatment . . . . .	6
1.3 Problems & Challenges in diagnosis . . . . .	8
1.3.1 Digital Pathology: Current status . . . . .	9
1.4 Deep Learning . . . . .	10
1.5 Goal of Thesis . . . . .	15
2. CLASSICAL APPROACHES . . . . .	17
2.1 Image analysis based methods . . . . .	17
2.1.1 Color normalization . . . . .	17
2.1.2 Nuclear Segmentation . . . . .	19
2.1.3 Feature extraction . . . . .	19
2.2 Challenges . . . . .	21
3. DEEP LEARNING FOR DETECTING METASTATIC BREAST CANCER	23

3.1	Whole Slide Image classification pipeline . . . . .	23
3.1.1	Region of Interest (ROI) detection with Image processing . . . . .	23
3.1.2	Construct training data: Tiling ROI . . . . .	26
3.1.3	Deep ConvNet for tile-based classification . . . . .	29
3.1.4	Building Tumor probability heat-maps . . . . .	33
3.1.5	Post-processing on heat-maps for slide-based classification . . . . .	36
4.	EXPERIMENTAL RESULTS . . . . .	40
4.1	Dataset . . . . .	40
4.2	Experimental setup . . . . .	40
4.3	Evaluation metrics . . . . .	42
4.3.1	Receiver Operating Curves (ROC) . . . . .	42
4.4	Model D-1 results . . . . .	43
4.4.1	Training . . . . .	43
4.4.2	ROC . . . . .	44
4.5	Ensemble method . . . . .	44
4.5.1	ROC . . . . .	48
4.6	Further improvements . . . . .	48
4.6.1	Feature selection . . . . .	48
4.6.2	Use of better classifier . . . . .	49
4.6.3	ROC . . . . .	50
4.7	Results comparison . . . . .	50
5.	CONCLUSION AND FUTURE WORK . . . . .	52
	REFERENCES . . . . .	54
	BIOGRAPHICAL STATEMENT . . . . .	60

## LIST OF ILLUSTRATIONS

Figure		Page
1.1	Cancer cells reproduction . . . . .	2
1.2	Multi-resolution Whole Slide Image [42] . . . . .	9
1.3	A layer model depicting steps involved in preparation of WSI . . . . .	11
1.4	ConvNet Architecture . . . . .	13
1.5	Convolution operation . . . . .	14
1.6	Pooling operation . . . . .	15
1.7	WSI Classification System . . . . .	16
2.1	Classical method architecture . . . . .	18
2.2	An example of stain normalization on a heavily over stained source image (a) Source Image (b) Target Image (c) Output Image [2] . . . . .	18
2.3	An example of Nuclear Segmentation. <i>Left:</i> Source Image. <i>Right:</i> Segmented Image . . . . .	20
2.4	Summary of object-level features used in histopathology image analysis [22] . . . . .	21
2.5	Summary of topological features used in histopathology image analysis [22] . . . . .	22
3.1	Cancer metastases detection framework [1] . . . . .	24
3.2	Closing operation with 3x3 window size . . . . .	26
3.3	Opening operation with 3x3 window size . . . . .	27

3.4	ROI detection process. <i>Top left</i> : Original RGB image. <i>Top middle</i> : converted HSV image. <i>Top right</i> : Filtered mask. <i>Bottom left</i> : mask after closing operation. <i>Bottom middle</i> : mask after opening operation. <i>Bottom right</i> : original image with highlighted ROI using blue curves .	27
3.5	Patch extraction from WSI. Tissue and Tumor regions are highlighted with green and red curves respectively. Area within green region but outside of red region is considered as Normal. Extract tumor patches from tumor regions and normal patches from normal regions [1] . . . . .	28
3.6	Data augmentation techniques . . . . .	29
3.7	<i>Top</i> : Images showing detected mitosis. Detected mitosis are circled green (true positives) and red (false positives); cyan denotes mitosis which are not detected. <i>Bottom</i> : Mitosis detection framework [19] . . . . .	30
3.8	Model D-1 training details . . . . .	31
3.9	Training loss for Deep ConvNet model. Note significant decrease in training loss when learning rate decrease. Here 1 iteration corresponds to 1 batch of 32 images . . . . .	32
3.10	Inception-V3 architecture [9] . . . . .	34
3.11	Inception modules where each 5 x 5 convolution is replaced by two 3 x 3 convolution [10] . . . . .	35
3.12	Inception modules after the factorization of the n x n convolutions. We chose n = 7 for the 17 17 grid [10] . . . . .	35
3.13	Inception modules with expanded the filter bank outputs. This architecture is used on the coarsest ( 8 x 8 ) grids to promote high dimensional representations [10] . . . . .	35
3.14	Process of building tumor probability heat-map . . . . .	36
3.15	Process of extracting features from heat-map . . . . .	37

3.16	Feature importance map . . . . .	39
4.1	Visualizing metastases using ASAP [42] . . . . .	41
4.2	ROC curve for Model D-1 predictions on Test dataset . . . . .	44
4.3	Model D-1 false positives. <i>Left</i> : Original slide. <i>Middle</i> : Ground truth. White pixels inside ROI represents ground truth. <i>Right</i> : model D-1 heat-map. Red pixels inside highlighted yellow regions represent false positives . . . . .	45
4.4	Example for preparing ensemble heat-map for slide Tumor_074. <i>Top</i> : Model D-1 heat-map. <i>Bottom</i> : Ensemble heat-map . . . . .	47
4.5	ROC curve for ensemble method on Test dataset . . . . .	48
4.6	Final ROC curve on Test dataset, obtained after applying feature se- lection and better classifier . . . . .	50
4.7	ROC comparison with Camelyon16 Top-5 methods . . . . .	51

## LIST OF TABLES

Table	Page
3.1 Inception-V3 outline [10] . . . . .	33
4.1 Number of slides in the Camelyon'16 dataset . . . . .	42

## CHAPTER 1

### INTRODUCTION

#### 1.1 Cancer

Cancer <sup>1</sup> is a non-communicable group of diseases that involves the abnormal growth of cells in a manner that it takes over normal cells and that the body cannot control it. It becomes very hard for the human body to function normally when this happens. Cancer can be treated very effectively, also, it is known that people lead much better and richer lives after cancer treatment. Cancer can be of various types, depending on the body part it has affected. Cancerous cells can be found in the colon, the breast, the lungs, and also in the blood of a patient. All cancers are similar - they usually differ in the way they grow and spread. To explain cancer, we need to understand what cells are and how they function. The cell is the basic structural, functional and biological unit of all living organisms. Each cell has a task to perform. There are cases in which a cell is damaged, or it may get worn out. In this case, the cell is killed and is replaced with a new healthy cell. One of the most important attributes of a cell is the natural manner of its reproduction. A healthy cell divides in an orderly manner, and is killed when they are worn out or damaged. Once these cells are killed, the new cells come in and take their place. Cancer is when the growth of these damaged cells cannot be controlled. During a cancer, these damaged cells keep growing and making new cells, effectively crowding out healthy cells, which causes a lot of problems in the organ the cancerous cell originated from. Cancer cells can also travel to other parts of the body, like for instance, cancer cells can travel from

---

<sup>1</sup><http://www.cancer.org/cancer/>

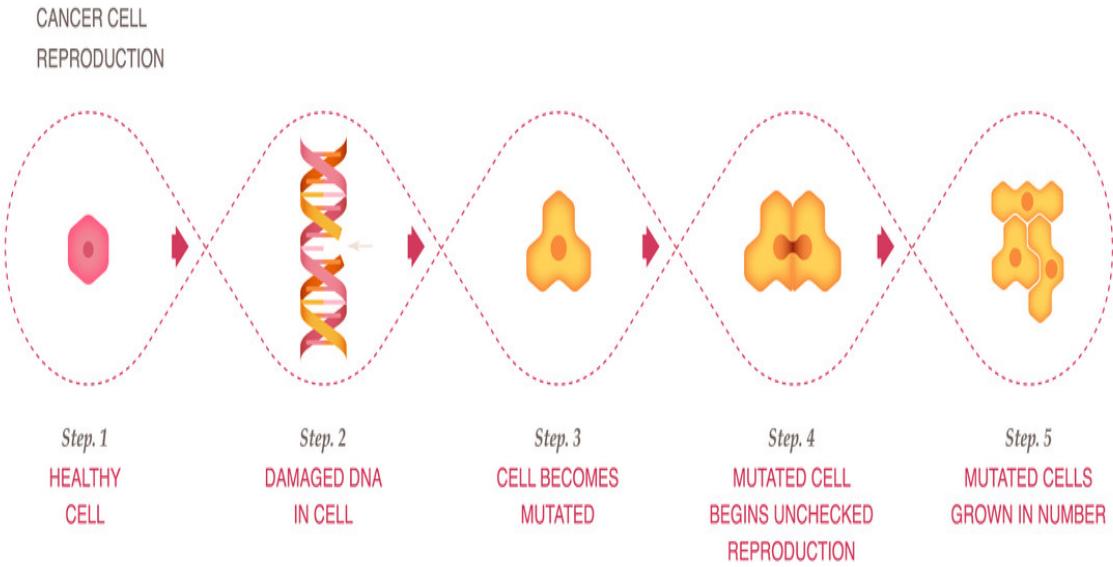


Figure 1.1: Cancer cells reproduction

the lung to the bones and grow there. This spreading of cancer cells is known as metastasis. Even though cancer can spread to other organs, it is classified based on the organ of origin. Hence, even when lung cancer spreads to the bones, it is still called lung cancer. To a doctor, cancerous bone cells look just like cancerous lung cells, but cancerous bone cells are classified as bone cancer only if it originated in the bone. Some cancers grow and spread rapidly, whereas others grow slowly. Different cancers respond to treatment in different manners. Cancer is mostly treated with surgery, but some types of cancer are treated with a treatment method known as chemotherapy. To get the best results in treatment, two or more treatment methods are generally used. When a patient is diagnosed with cancer, the doctor will want to find out what kind of cancer the patient is suffering with, in order to choose an effective treatment plan. This is to ensure that the right treatment is chosen to help treat the patient's cancer. Most cancers are known to form a lump called a tumor or a growth, but not all lumps are cancerous. A cancerous lump is a malignant lump,

whereas a non cancerous lump is a benign lump. But not all cancers form lumps. Cancers like leukemia don't form tumors. They grow either in blood cells, or in other cells of the body. To help choose the patient's treatment plan, the doctor also needs to have information about how far the cancer has spread from it's origin. This is known as the cancer stage. The farther the cancer has travelled from it's origin, the higher the stage the cancer of the patient is, which could be either in stage 1,2,3 or 4. The knowledge of this information helps the doctor decide which treatment to choose. A lower stage cancer, like stage 1 or 2, means that the cancer has not spread very much, whereas a higher stage like stage 3 or 4 means that it has spread much more. Stage 4 is the highest stage of cancer that can be diagnosed in a patient.

## 1.2 Breast cancer

Breast <sup>2</sup> cancer is a cancer that develops from breast tissue. It is said to be the second leading cause of death among women. It is also estimated that over 40,000 women diagnosed with breast cancer die every year in the United States [46]. The cancer starts in the cells of the lobules, which are the milk producing glands, or the ducts, the passages that drain milk from the lobules to the nipple. This group of cancer cells can then invade surrounding tissues or spread (metastasize) to other areas of the body, by entering blood cells or lymph vessels that branch to all parts of the body. This process of the cancerous cells travelling to all parts of the body is known as metastasis.

### 1.2.1 Types of Breast cancer

1. **Ductal Carcinoma in Situ:** Ductal Carcinoma in Situ (DCIS) is a non-invasive breast cancer where abnormal cells have been contained in the lining

---

<sup>2</sup><http://www.nationalbreastcancer.org/about-breast-cancer>

of the breast milk duct. Carcinoma in Situ is used to describe early stage of cancers. Carcinoma means “cancer” and in situ means “in the original place.” In this type of breast cancer, the atypical cells have not spread outside of the ducts into the surrounding breast tissue. This type of cancer in the early stages is highly treatable, however, if left untreated or undetected, can spread into the surrounding breast tissue. This type of breast cancer is classified under stage 0.

2. **Invasive Ductal Carcinoma:** Invasive Ductal Carcinoma means that abnormal cells that originated in the lining of the breast milk duct have spread beyond the ducts and invaded surrounding tissue. It is the most common type of breast cancer, making up nearly 70-80% of all breast cancer diagnoses, and is known to also commonly affect men. The stage of this breast cancer varies on a case to case basis.
3. **Triple Negative Breast Cancer:** Triple negative breast cancer means that the cells in the tumor are negative for progesterone, estrogen, and HER2/neu receptors, which are receptors known to fuel most breast cancer growth. Chemotherapy is an effective option to treat this kind of breast cancer, as common treatments such as hormone therapy and drugs that target these receptors become ineffective. Also, 10-20% of breast cancer is known to be triple negative. This type of cancer is most likely to affect younger people, African Americans, Hispanics, and/or patients with a BRCA1 gene mutation.
4. **Inflammatory Breast Cancer:** Inflammatory breast cancer is a less common form of breast cancer that produces no distinct tumor or hump and is isolated within the breast. It is an aggressive and fast growing breast cancer in which cancer cells infiltrate the skin and lymph vessels of the breast. Symptoms begin to appear when the lymph vessels become blocked by the breast cancer cells.

This type of breast cancer is classified as a stage 3 breast cancer and requires aggressive treatment.

5. **Metastatic Breast Cancer:** Metastatic breast cancer is cancer that has spread beyond the breast, into other parts of the body such as the lungs, liver, bones, or brain. This cancer is classified as a stage 4 cancer, and is generally incurable. The symptoms of this cancer vary based on where it has spread, and it can generally spread to the bones, brain, liver and lungs.

### 1.2.2 Diagnosis

1. **Mammogram:** A mammogram is an x-ray of the breast. This allows a qualified specialist to examine the breast tissue for any signs of breast cancer. Screening mammograms are routinely administered to women with no apparent symptoms, to detect whether they have any cancerous cells. Diagnostic mammograms are taken when any abnormalities are found in the screening mammogram. Some abnormalities may include a lump, breast pain, nipple discharge, thickening of skin on the breast and changes in the size or shape of the breast. In a diagnostic mammogram, more x-rays are taken, providing views of the breast from multiple vantage points, to provide a more detailed x-ray of the breast.
2. **Ultrasound:** A breast ultrasound is recommended when a suspicious site is found on a patient's screening mammogram. It is a scan that uses penetrating sound waves that do not affect or damage the tissue and cannot be heard by humans. The breast tissue deflects these waves, and the computer then uses these deflections to paint a picture of what is going on inside the breast tissue.
3. **MRI:** MRI is a radio imaging technique that could also be used to get an understanding of what's going on inside a patient's breast. During a breast MRI, a magnet connected to a computer transmits magnetic energy and radio

waves (not radiation) through the breast tissue. It scans the tissue, and makes detailed pictures of areas that are within the breast, to help the diagnosing physician distinguish between a normal and diseased tissue.

4. **Biopsy:** A breast biopsy is the only diagnostic procedure that can determine if the suspected area is cancerous. It is a test that removes tissue or sometimes fluid from the suspicious area. The removed cells are examined under a microscope and further tested to check for the presence of breast cancer. Biopsies are of three types - Fine needle aspiration, core-needle and surgical biopsy. Core-needle and surgical biopsy are commonly used on the breast. Factors such as appearance, size and location of the suspicious area of the breast help a doctor decide the type of biopsy to recommend.

#### 1.2.3 Treatment

It is very important for the patient to have a positive relationship with the doctor. The patient should work with the doctor to see what options they have, for treatment of their cancer. It is also important for the patient to understand the difference between going through a standard treatment and clinical trial treatment, so that the patient can make an informed decision when it comes to their treatment choice. Breast cancer standard treatments are those treatments that are recommended by experts, and what experts agree are appropriate, accepted and widely used. These are standard procedures that have been tried and tested, and have proved useful in fighting breast cancer in the past. A breast cancer clinical trial is an approved research study in which a patient goes through treatments that differ from the standard ones. Some doctors believe that these new treatments have a potential to some day replace the existing standard treatments, and become new standards in treating breast cancer. When the treatments administered in clinical trials can prove to

perform better than the existing treatments, they then become the standard, and are then administered during standard treatments. Hence, all current standards were clinical trials at one time.

Surgery, chemotherapy and radiation are the most common treatments of cancer. Surgery is one of the oldest types of cancer therapy. It is the procedure done on the patient where the tumor is removed, along with any other surrounding tissue that is affected by it. This is done to take out the cancerous tissue from the body, to prevent it from spreading. Chemotherapy is the method of treatment where drugs are used to cure the cancerous patient. The drug either kills the cancerous cells, or slows its growth. Radiation therapy is when the patient is exposed to high density waves or particles, such as x-rays, gamma rays or electron beams to kill the cancerous growth.

Some of the cancers where a surgery is performed is on breast cancer and prostate cancer. For breast cancer, part or all of the breast may be removed, depending on the stage of cancer the patient is in. An early stage patient can have a breast conserving surgery where only the cancerous tissue is removed, but a later stage patient has to go for a mastectomy, where the entire breast is removed, sometimes along with nearby tissues. For prostate cancer, the prostate gland is removed during surgery, the main type of which is called a radical prostatectomy.

Chemotherapy (chemo in short) is used to treat blood related cancers like leukemia. The type of treatment given to a patient suffering with leukemia depends on the type of leukemia the patient suffers from. Leukemia is classified based on the type of blood that is cancerous in the patient, either the Red Blood Cells, the White Blood cells [16], or the platelets. It is also classified based on its rate of growth, which is either acute for fast growing cancer, and chronic for a slow growing cancer. Also, attributes such as the age of the patient, whether the cancer has spread to the brain

or spinal cord, whether there are certain changes in the genes and whether the cancer has been treated before also dictate the type of chemotherapy treatment needed to be given to the patient. Some chemo can be given using IV (into a vein through a needle), and others are in form of pills.

### 1.3 Problems & Challenges in diagnosis

Pathology is the medical field that primarily deals with diagnosis and treatment of disease. This is the field of medicine that is given the responsibility to provide a subjective diagnosis, to guide a patient with their treatment, and to provide management decisions to the patient [4]. The recent advents in computing pathology also made it possible to predict survival of a cancer patient [30, 32]. It is very important that the field of precision medicine make constant advances. This is to ensure an accurate diagnosis of the cancer that the patient is suffering from. It is important that there be standardized, accurate and reproducible pathological diagnoses for advancing precision medicine. As stated by [1], if we look behind at the past, the microscope was the primary tool used by pathologists, ever since the mid 19th century. The images formed by these microscopes had many limitations after qualitative visual analysis of them, which included lack of standardization, diagnostic errors and the significant cognitive load required to manually evaluate millions of cells across hundreds of slides. Let us consider the evaluation of the breast sentinel lymph nodes, as it is considered today as a very important component. Patients with a sentinel lymph node positive for metastatic cancer frequently results in more aggressive clinical management [38, 39]. To manually conduct a pathological review of the sentinel lymph nodes is very time consuming and laborious, especially in cases where the lymph nodes are negative or contain very small cancerous cells [27]. To improve accuracy of metastatic detection, many clinical laboratories have tried applying proteins such as immuno-

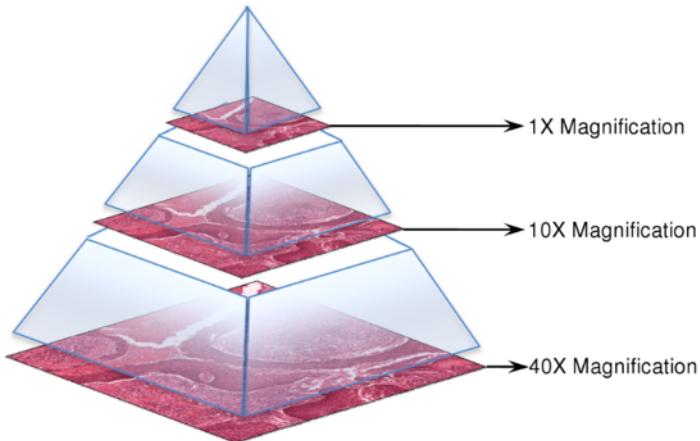


Figure 1.2: Multi-resolution Whole Slide Image [42]

histochemistry [25] for pancytokeratins [5] on breast cancer cells; however, there are many limitations to pancytokeratin immunohistochemistry testing of sentinel lymph nodes which include increased cost, increased time for slide preparation, increased number of slides required for pathological review, and less accuracy [1].

### 1.3.1 Digital Pathology: Current status

Pathology is a medical field that is over a 150 years old, which has progressed by leaps and bounds thanks to the advent of digital pathology [2]. Virtual microscopy is partially credited for the existence of digital pathology, which is the practice of converting glass slides into digital slides that can be viewed, managed, and analyzed on a computer monitor [41]. Thanks to the progress in the field of Whole-Slide Imaging (WSI), the digital pathology field has exploded and has been named as one of the most promising avenues of diagnostic medicine in order to achieve even better, faster and cheaper diagnosis, prognosis and prediction of cancer and other important diseases [20]. Figure 1.3<sup>3</sup> displays a layer model depicting steps involved

---

<sup>3</sup>[http://www.hopkinsmedicine.org/mcp/PHENOCORE/CoursePDFs/2013/13\\_19\\_Cornish\\_Digital\\_Path.pdf](http://www.hopkinsmedicine.org/mcp/PHENOCORE/CoursePDFs/2013/13_19_Cornish_Digital_Path.pdf)

in preparation and viewing of WSI. As stated by [2], though adoption of digital pathology has increased in various medical centers, the industry has not yet entered into clinical diagnostics due to inherent problems, such as human variability from tissue acquisition, improper staining techniques, and subjectivity in diagnosing under a microscope. A pathologist would then look for patterns in a tissue sample and use their medical training to interpret those patterns and make a diagnosis. As demonstrated by many computer applications in myriad industries, computer-assisted pattern recognition either supersede or are in par with a human’s ability to recognize patterns. Though digital pathology is on the peak of wide-spread adoption, the difficulties due to the variability of hardware scanning and fear of computation tools in which the user has no idea of the underlying source code, as known as “black-box” tools, has lead to a longer adoption curve than compared to other medical specialties that have gone totally digital, like radiology [2]. Over the past several decades there has been an interest in developing computer software to assist in the analysis of digital microscopic images in pathology. Therefore, computer-assisted image analysis systems have been developed to aid in the detection of metastatic tissues from digital slides of sentinel lymph nodes; however, clinically, these systems are not used due to the lack of standardization of image formats, system noise, and lack of clinical and technical studies on digital pathology systems [2]. Hence, active research is currently taking place to develop effective and cost efficient methods for sentinel lymph node evaluation, as there is a heavy requirement for a high-performing system that could increase accuracy and reduce cognitive load at low cost.

#### 1.4 Deep Learning

Deep Learning is a new area of Machine Learning research. It is a technique to use very deep (in terms of number of layers) neural network to solve problems related

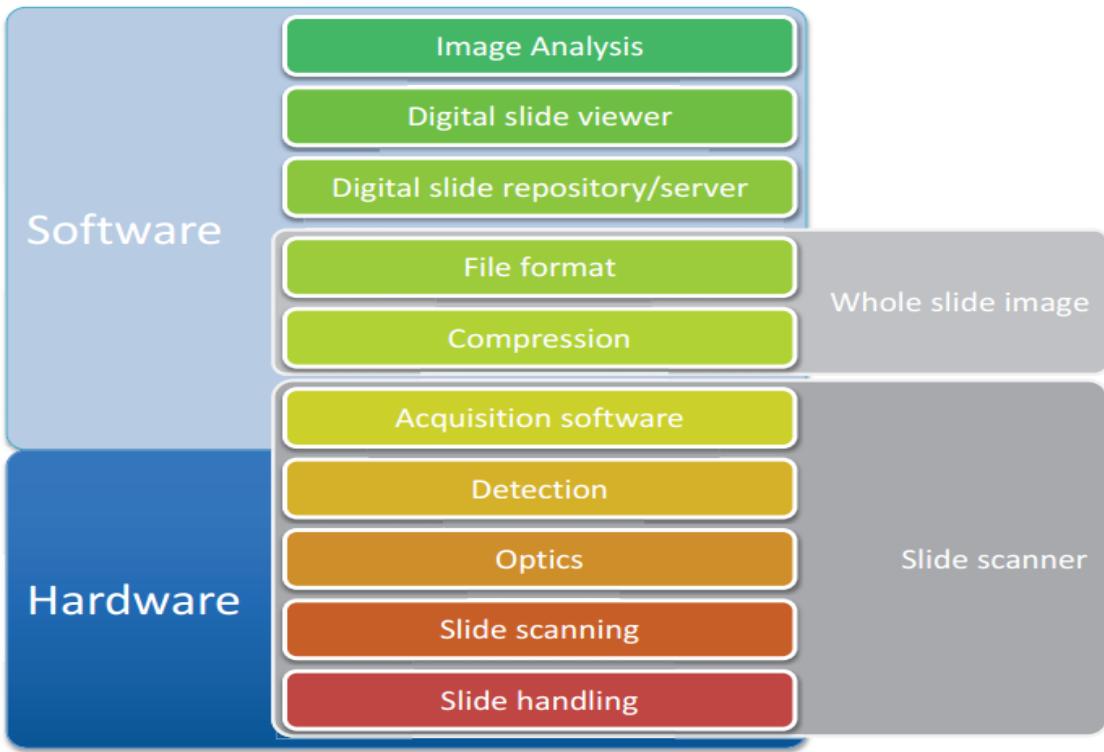


Figure 1.3: A layer model depicting steps involved in preparation of WSI

to visual recognition. It was introduced with the aim to bring Machine Learning closer to Artificial Intelligence. Until 1990, neural network was not getting acceptance into machine learning industry as there was no proper way to train a good network. But in 1990, the advent of back-propagation [24] revolutionized AI industry and suddenly neural network became the hot favorite topic in the field of machine learning research. There have been many problems in various fields of Computer Science that have been solved with the help of various Deep Learning architectures. Deep Learning architectures have been applied to problems in fields like computer vision [8, 9, 10, 11, 12, 13], automatic speech recognition [26], natural language processing, audio recognition and bio-informatics, using deep learning architectures such as deep neural networks, convolution deep neural networks, deep belief networks and recurrent

neural networks. The results obtained were state of the art for various tasks given to the network. The key aspect of all these deep learning architectures is the use of Convolutional Neural Network (ConvNet). ConvNet is a biologically inspired form of the artificial neural network, that has local connections and shared weights. It is one of the most important tools of machine learning when it comes to the current generation, and it has been very popularly used to solve image recognition tasks, in the field of Computer Vision. Some of the popular and highly used ConvNet models are <sup>4</sup>:

1. **LeNet** [7]: Not only is LeNet one of the first successful applications of Convolution Networks, it is also considered to be an excellent “first architecture” for ConvNets. Developed in 1990, the best known LeNet architecture is one that was used to read zip codes, digits, etc.
2. **AlexNet** [8]: AlexNet is the first work that popularized the use of Convolution Networks in Computer Vision. It is a large, deep convolution network used to classify over 1.3 million high resolution images of the LSVRC-2010 ImageNet training set into 1000 different classes. During the time AlexNet was developed, it was common to have only a single Convolution (CONV) layer and then to have it be immediately followed by a POOL layer, but the AlexNet was designed in a different manner. It had a very similar structure to LeNet, the difference being that it was deeper, bigger and featured Convolution layers stacked on top of each other.
3. **GoogLeNet** [9]: Developed by Google, GoogLeNet is the winning deep convolution network architecture from the ImageNet Large-Scale Visual Recognition Challenge 2014 (ILSVRC 2014). This winning architecture was code named “Inception”. Some important features of this architecture was the development

---

<sup>4</sup><http://cs231n.github.io/convolutional-networks/>

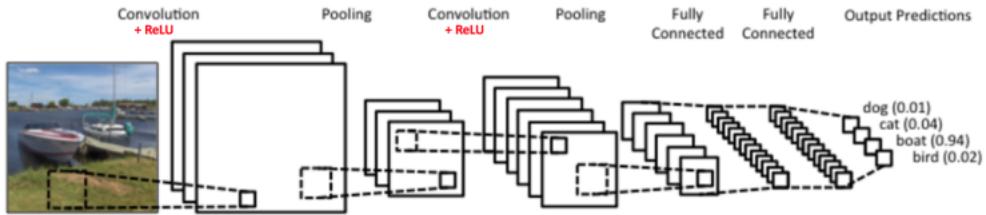


Figure 1.4: ConvNet Architecture

of an Inception module to dramatically reduced the number of parameters in the network , from 60M as in AlexNet to 4M. Another interesting approach to this model was to use Average Pooling instead of Fully Connected (FC) layers after Convolution layer, which helps to eliminating significant number of parameters that are not important for classification.

4. **VGGNet** [11]: VGGNet was declared the runner-up architecture in the ImageNet Large-Scale Visual Recognition Challenge 2014 (ILSVRC 2014). The main aspect of this network was that it showed that the depth of the network is a critical component for a well performing neural network. Their final best network contains 16 CONV/FC layers. This network also features an extremely homogeneous architecture that only performs 3x3 convolutions and 2x2 pooling from the beginning to the end.
5. **ResNet** [12]: ResNet (Residual Network) is the winning architecture in the ImageNet Large-Scale Visual Recognition Challenge 2015 (ILSVRC 2015). The two major features of this architecture are skip connections and a heavy use of batch normalization. This architecture has missing fully connected layers at the end of the network. By far, ResNets are currently state-of-the-art Convolutional Neural Network models and are the default choice for using ConvNets in practice.

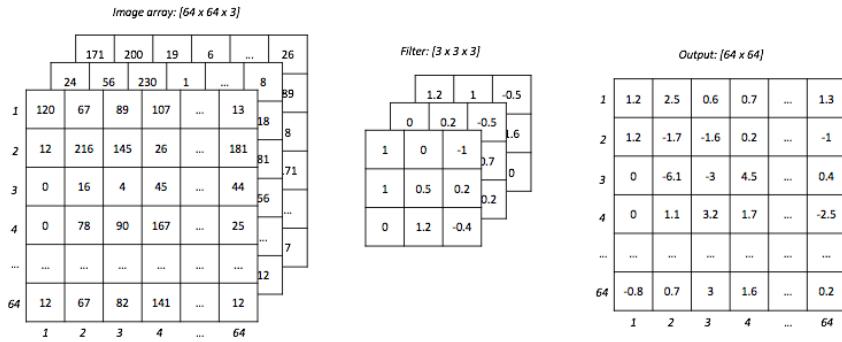


Figure 1.5: Convolution operation

Figure 1.4<sup>5</sup> illustrates an architecture for ConvNet. There are three main types of layers that are developed to build a ConvNet: Convolutional Layer, Pooling Layer, and Fully-Connected Layer. The functions of each of the components in a convolution network is as follows:

1. **CONVOLUTION:** This is the layer that computes the output of neurons that are connected to the local receptive field of the input. This computes a dot product between weights and local region of input volume. This is illustrated in Figure 1.5.
2. **POOLING:** The pooling layer performs the down sampling operation along the spatial dimensions of the input, as shown in Figure 1.6.
3. **FULLY CONNECTED:** This layer is similar to a neural network, wherein each neuron is connected to all the neurons in previous layers. This layer helps to estimate class scores.

---

<sup>5</sup><https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/>

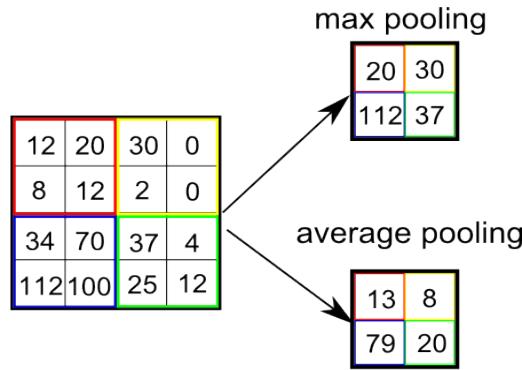


Figure 1.6: Pooling operation

## 1.5 Goal of Thesis

The goal of this thesis is to develop deep learning-based classification pipeline for detection of cancer metastases from whole slide images of breast sentinel lymph node. The classification pipeline consists of five different stages:

1. Region of Interest (ROI) detection with Image processing
2. Construct training data: Extract Positive & Negative tiles from ROI
3. Train Deep ConvNet model for tile-based classification
4. Building tumor probability heat-maps using trained model
5. Post-processing on heat-maps for slide-based classification

Traditional computer aided methods are not useful for clinical practices because they require pathologist to set several manual parameters to obtain accurate results hence proves burdensome. Thus the key aspect of this thesis is to develop proposed method in a way that it will prove highly useful and can be easily operated in clinical settings with minimal human intervention. Evaluate effectiveness of the proposed system by performing extensive experiments on real life breast cancer data-set available as part

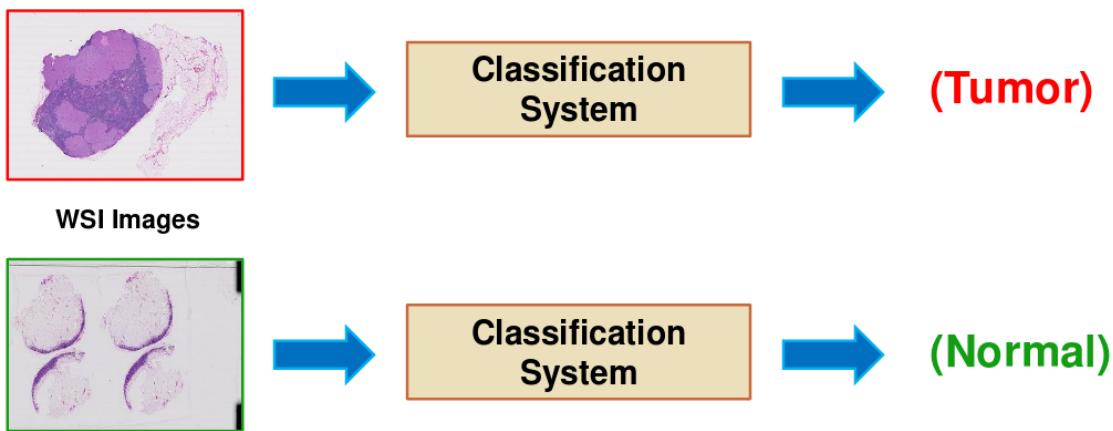


Figure 1.7: WSI Classification System

of Camelyon'16<sup>6</sup> grand challenge. Figure 1.7 depicts an overview of the proposed whole slide image classification system for identifying metastatic cancer.

---

<sup>6</sup><https://camelyon16.grand-challenge.org/>

## CHAPTER 2

### CLASSICAL APPROACHES

#### 2.1 Image analysis based methods

Since a long time, there have been an interest of scientists to develop computer aided methods for identifying cancer metastasis from GigaPixel ( $10^6 \times 10^6$  pixels) histological images. These historical methods are mainly focused on low level image analysis [33] tasks such as color normalization, nuclear segmentation, and feature extraction [21, 22]. Detailed architecture of these methods is illustrated in Figure 2.1. In this section, we will explain each stage of the classical method in detail.

##### 2.1.1 Color normalization

Color normalization is a technique used for reducing the difference between different tissue samples, introduced due to applied staining and various scanning conditions during preparation of whole slide images [15]. With staining, it becomes easier to analyze images as it helps to highlight structural elements into whole slide images. According to [22], there are multiples ways in which staining can be applied. One method is to calibrate targets or finding pixel intensity patterns from multiple images and then fit polynomial surface over source images. Another highly used method in industry is to match histograms of different source images. [22] also mentioned that, approach based on gradient calculation from LUV color space proved reliable for highlighting tissue structures from WSI images. As color normalization helps in building generalized solution, it is considered to be the most important step for build-

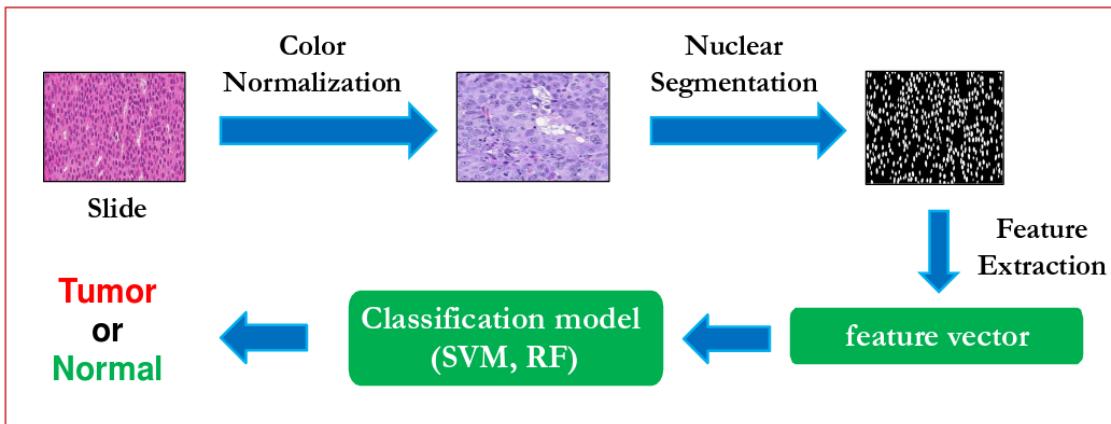


Figure 2.1: Classical method architecture

ing histological image analysis technique. Figure 2.2 illustrates an example of Color Normalization operation on digital WSI.

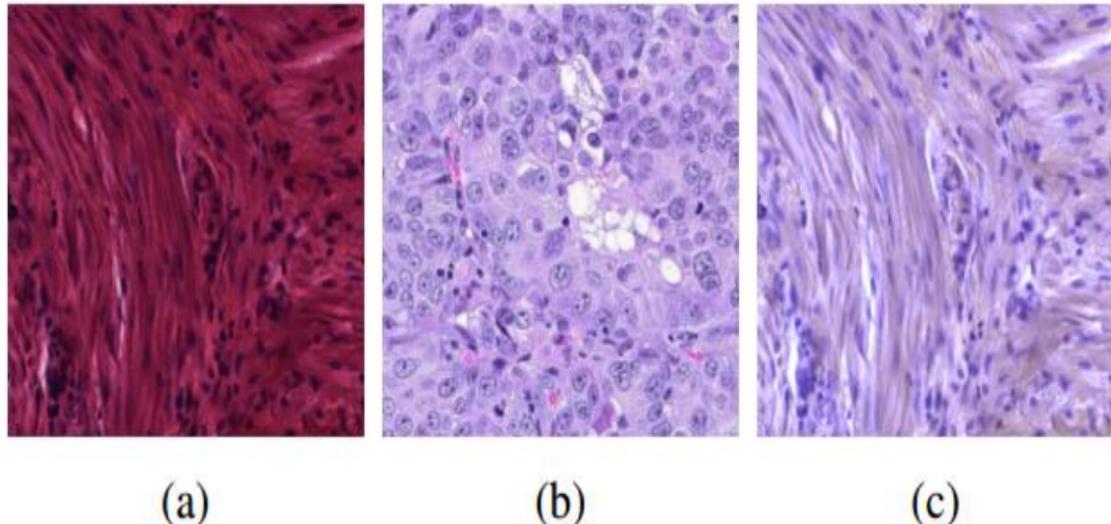


Figure 2.2: An example of stain normalization on a heavily over stained source image  
(a) Source Image (b) Target Image (c) Output Image [2]

### 2.1.2 Nuclear Segmentation

The second step after color normalization is to segment nuclei from normalized images. Nuclear segmentation is another important step in histological image analysis. It is the process of identifying nuclei structure present in whole slide images [28]. It is very important to determine size and shape of nuclear structures as these properties help pathologist to understand the seriousness of disease if present [22]. More over, these properties also allows to determine sub-types [29] of detected cells which helps pathologist to determine specific type of cancer disease patient is suffering from. For example, the presence of large number of malignant cells in a breast cancer histological image is a clear indication of a higher cancer stage. As per [22], other significant measure to decide severity of disease is to count mitosis from histological images as higher mitosis count is a strong indication of poor disease outcome for some cancer situations. Due to its high importance, it remains a great interest of scientist to develop automated methods for identifying nuclear structures [31] from histological images. Two of the methods developed are [22]:

1. **Global segmentation:** Here entire image is processed in a single pass to simultaneously detect all the nuclei from an image.
2. **Local segmentation:** In contrast to global segmentation, here only a portion of the image is processed at a time to identify specific structure.

Figure 2.3 illustrates an example of Nuclear Segmentation operation on digital WSI.

### 2.1.3 Feature extraction

The third stage in classical image analysis method is to extract object level and topological features from the nuclear-segmented image. Object level features correspond to physical appearance of cell nuclei while topological features are related to inter cell structure being formed by the arrangement of multiple cells in a particular

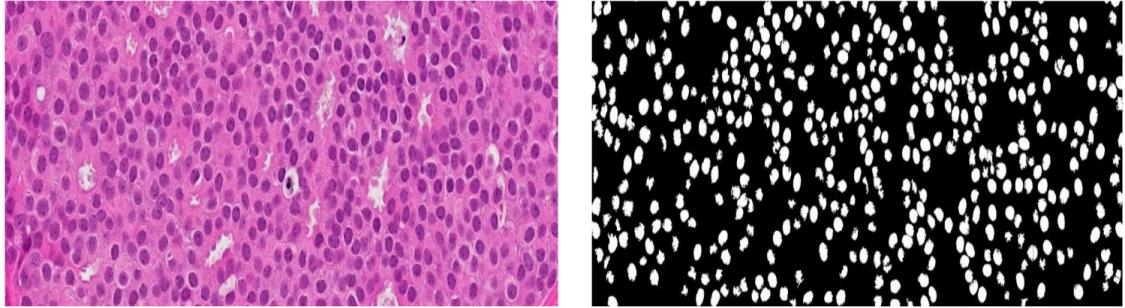


Figure 2.3: An example of Nuclear Segmentation. *Left*: Source Image. *Right*: Segmented Image

region [6]. Object level features can be separated into categories such as shape, size, radiometric & densitometric, texture, and chromatin-specific [22]. Here size and shape are real object level features while others are considered to be low level features. The example of topological features includes graph based properties such as minimal spanning tree, connected components, k-NN [34, 35, 36, 37] graph etc. Topological features are used to formulate tissue states and also to compare different tissues by formulating matrices from these graphs and classifying values. Summaries of object-level and topological features are listed in Figure 2.4 & Figure 2.5 respectively. In addition to object and topological features various statistic features such as mean, median, minimum, maximum, standard deviation, skewness, and kurtosis can also be used for classification purpose. Normally, for binary image number of features are in order of 100 but for RGB image these features are extracted for each R, G & B channel separately, hence the number of features can easily exceed 1000 [22]. In classical approach, feature extraction process is typically followed by final classification stage where classification models such as Support Vector Machine (SVM) and Random Forest (RF) are developed to classify these features into two or more categories based on the nature of the problem under consideration [43, 44, 45].

Category	Features
<b>Size and Shape</b>	Area
	Elliptical Features: Major and minor axis length, eccentricity, orientation, elliptical deviation
	Convex Hull Features: Convex area, convex deficiency, solidity
	Filled Image Features: Filled area, Euler number
	Bounding Box Features: Extent, aspect ratio
	Boundary Features: Perimeter, radii, perimeter Fourier energies, perimeter curvature, bending energy, perimeter fractal dimension
	Other Shape Features: Equivalent diameter, sphericity, compactness, inertia shape
	Center of Mass
	Reflection Symmetry
<b>Radiometric and Densitometric</b>	Image Bands, Intensity
	Optical density, integrated optical density, and mean optical
	Hue
<b>Texture</b>	Co-occurrence Matrix Features: Inertia, energy, entropy, homogeneity, maximum probability, cluster shade, cluster
	Fractal Dimension
	Run-length Features: Short runs emphasis, long runs emphasis, gray-level non-uniformity, run-length non-uniformity, runs percentage, low gray-level runs emphasis, high gray-level runs
	Wavelet Features: Energies of detail and low resolution images
	Entropy
<b>Chromatin-Specific</b>	Area, integrated optical density, mean optical density, number of regions, compactness, distance, center of mass

Figure 2.4: Summary of object-level features used in histopathology image analysis [22]

## 2.2 Challenges

Though classical methods are faster and cheaper in comparison to manual methods, they are not used in clinical practices due to various challenges induced by these methods. They completely rely on prior information such as size and shape of cells. Also they use hand crafted features which do not perform well in slide based classification. Normally cells have variety of different sizes (from tiny to very large) & shapes and due to which it is impossible to develop single generalized method for slide based classification which performs well for every type of cell structure. Moreover,

Graph Structure	Features
<b>Voronoi Tessellation</b>	Number of nodes, number of edges, cyclomatic number, number of triangles, number of k-walks, spectral radius, eigenexponent, Randic index, area, roundness factor, area disorder, roundness factor homogeneity
<b>Delaunay Triangulation</b>	Number of nodes, edge length, degree, number of edges, cyclomatic number, number of triangles, number of k-walks, spectral radius, eigenexponent, Wiener index, eccentricity, Randic index, fractal dimension
<b>Minimum Spanning Tree</b>	Number of nodes, edge length, degree, number of neighbors, Wiener index, eccentricity, Randic index, Balaban index, fractal dimension
<b>O'Callaghan Neighborhood Graph</b>	Number of nodes, number of edges, cyclomatic number, number of neighbors, number of triangles, number of k-walks, spectral radius, eigenexponent, Randic index, fractal dimension
<b>Connected Graph</b>	Number of nodes, edge length, number of triangles, number of k-walks, spectral radius, eigenexponent, Wiener index, eccentricity, Randic index, fractal dimension
<b>Relative Neighbor Graph</b>	Number of nodes, number of edges, cyclomatic number, number of neighbors, number of triangles, number of k-walks, spectral radius, eigenexponent, Randic index, fractal dimension
<b>k-NN Graph</b>	Number of nodes, edge length, degree, number of triangles, number of k-walks, spectral radius, eigenexponent, Wiener index, eccentricity, Randic index, fractal dimension

Figure 2.5: Summary of topological features used in histopathology image analysis [22]

these methods require several manual parameters to be set in order to obtain accurate results, thus proves burdensome for pathologists.

## CHAPTER 3

### DEEP LEARNING FOR DETECTING METASTATIC BREAST CANCER

#### 3.1 Whole Slide Image classification pipeline

In this thesis we tried to overcome the challenges faced by classical methods with the development of Deep Learning based classification pipeline for identifying metastatic breast cancer from digital whole slide images. Our classification pipeline consists of five stages:

1. Region of Interest (ROI) detection with Image processing
2. Construct training data: Extract Positive & Negative tiles from ROI
3. Train Deep ConvNet model for tile-based classification
4. Building tumor probability heat-maps using trained model
5. Post-processing on heat-maps for slide-based classification

Figure 3.1 depicts our cancer metastases detection framework.

##### 3.1.1 Region of Interest (ROI) detection with Image processing

The first stage in our pipeline is to identify tissue region (foreground) from whole slide image by excluding background white space. As described before, WSIs are very large GigaPixel ( $10^6 \times 10^6$  pixels) images thus processing even a single image takes a significant amount of time [3]. Finding ROI is an essential step as it helps to reduce computation time to a great extent by allowing to process only the region where probability of having tumor is more likely [1]. Similar to [1], we were able to remove approximately 82% of background region per WSI. Python open-CV APIs are

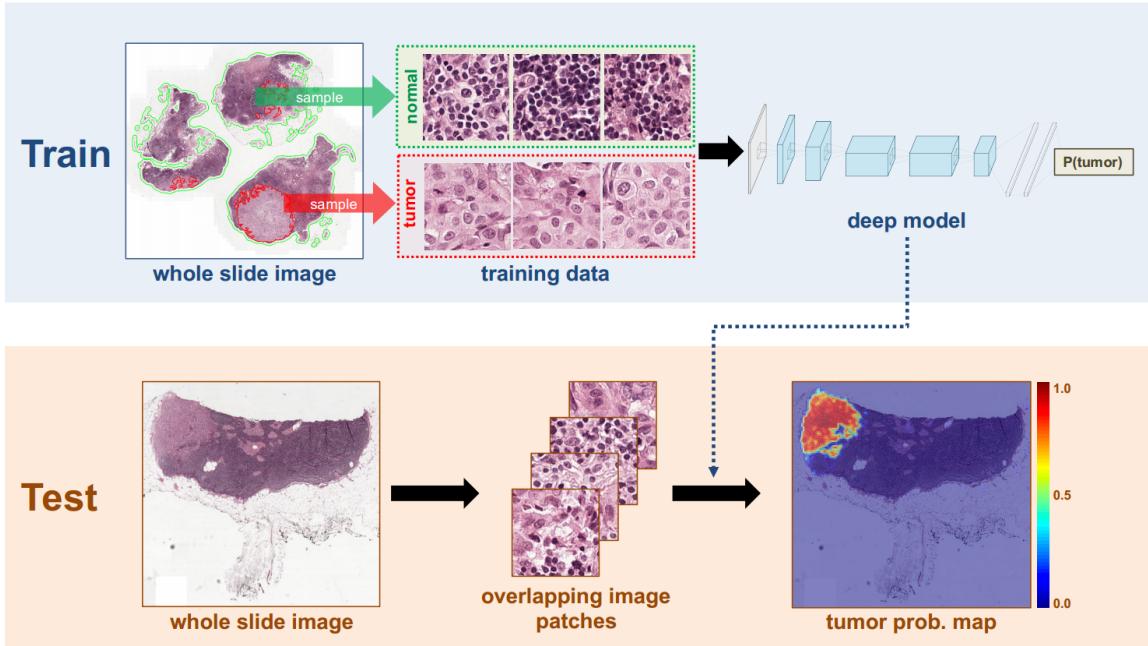


Figure 3.1: Cancer metastases detection framework [1]

used to perform various operations involved in finding ROIs. Our process includes five steps:

- 1. RGB to HSV conversion:** First step is to convert the original image from RGB color space to HSV (hue, saturation and value) color space. In HSV color space, analyzing color values is more convenient as values are more intuitive and easy to represent.
- 2. Binary mask generation:** Next step is to build a binary mask by filtering H, S & V component values in particular range. Specifically, we filtered pixels having values in the range 20 to 200 of H, S & V components. In industry, Otsu's [14] threshold based technique is quite popular for generating binary masks but we have developed our own method based on filtering pixel values as our method produces better results compared to Otsu. The binary mask

contains white pixels in areas where pixel values falls within the filtered range and black pixels everywhere else.

3. **Closing:** On a binary mask we apply closing operation to fill in small black color holes. Closing is defined simply as a dilation followed by an erosion using the same window size for both operations. In simple words, a fix sized window (e.g. 3x3) slides over the whole binary image and fill in all pixel in the current window if at least a single pixel in that window is found white. The effect of the operator is to preserve background regions that have a similar shape to a window, or that can completely contain the window, while eliminating all other regions of background pixels. Window size in closing operation plays an important role in producing effective results. Large window size will produce unwanted foreground noise while too small window size will leave many empty holes. Thus one has to be very careful while selecting window size in order to produce accurate results. After extensive experiments with different window sizes, we found that window size of 20 x 20 produces effective results for our dataset. Figure 3.2 <sup>1</sup> demonstrates the closing operation.
4. **Opening:** After closing we apply opening operation on a binary mask to eliminate small clumps of undesirable foreground pixels. It is basically a complete opposite of closing defined simply as an erosion followed by a dilation using the same window size for both operations. Like closing, in opening too a fix sized window (e.g. 3x3) slides over the whole binary image but here instead of fill in it erodes all the white pixel in the current window if a single pixel in that window is found black. As in closing, here also selecting proper window size is

---

<sup>1</sup><http://homepages.inf.ed.ac.uk/rbf/HIPR2/close.htm>

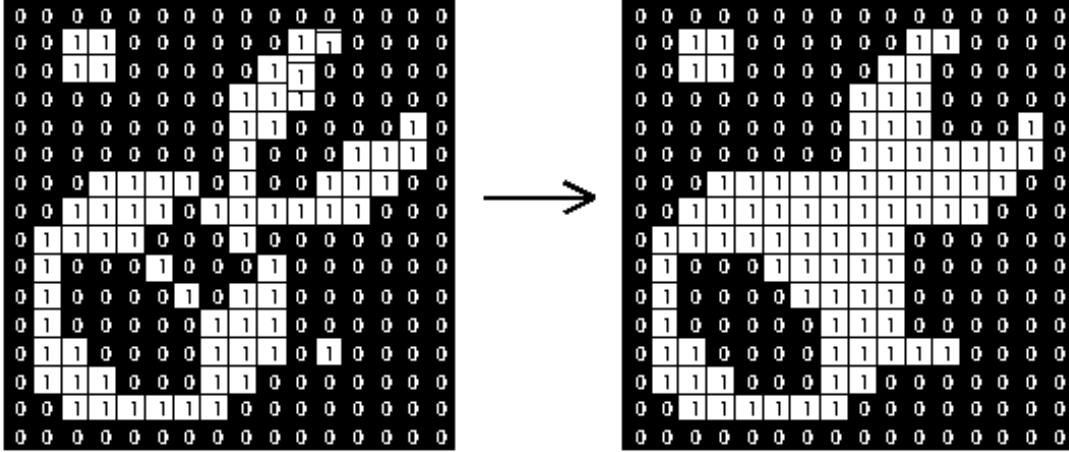


Figure 3.2: Closing operation with 3x3 window size

very important. Our experiments show that window size of 5 x 5 works best for our dataset. Figure 3.3<sup>2</sup> demonstrates the opening operation.

**5. Finding contours:** The last operation is to find contours from the binary mask obtained after opening operation. This operation derives the boundaries of white areas in the binary mask. After finding contours we draw them onto the original RGB image in order to highlight ROIs.

The whole process of ROI detection and final results are visualized in Figure 3.4.

### 3.1.2 Construct training data: Tiling ROI

This section explains the details about construction of training set, to be used for training Deep model. For this thesis we have used breast cancer dataset provide as a part of Camelyon'16 [42] grand challenge. Dataset consists of total 270 training and 130 test whole slide images. Training dataset includes 160 Normal and 110

---

<sup>2</sup><http://homepages.inf.ed.ac.uk/rbf/HIPR2/open.htm>

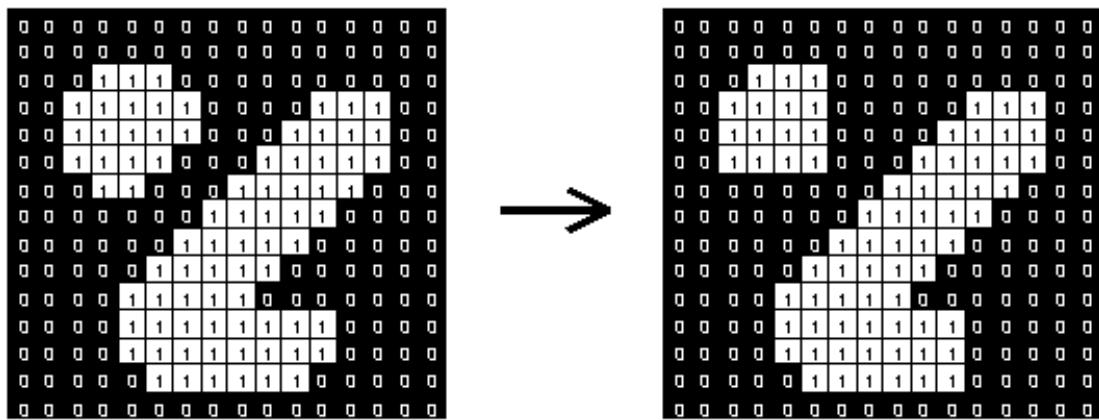


Figure 3.3: Opening operation with 3x3 window size

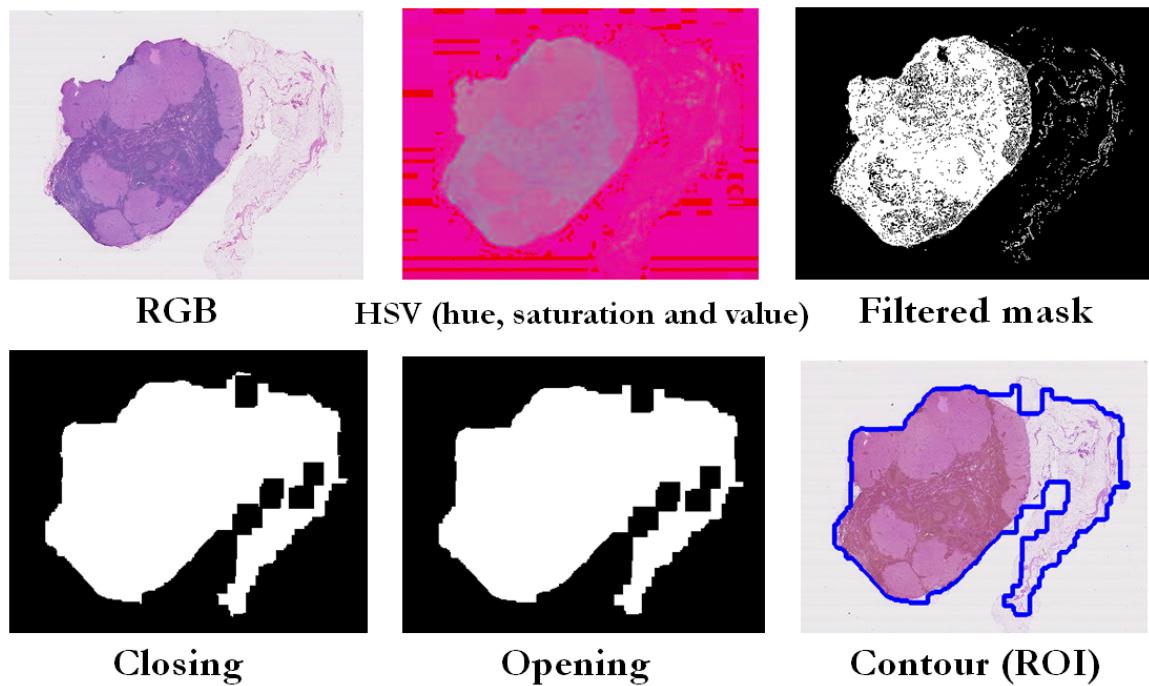


Figure 3.4: ROI detection process. *Top left:* Original RGB image. *Top middle:* converted HSV image. *Top right:* Filtered mask. *Bottom left:* mask after closing operation. *Bottom middle:* mask after opening operation. *Bottom right:* original image with highlighted ROI using blue curves

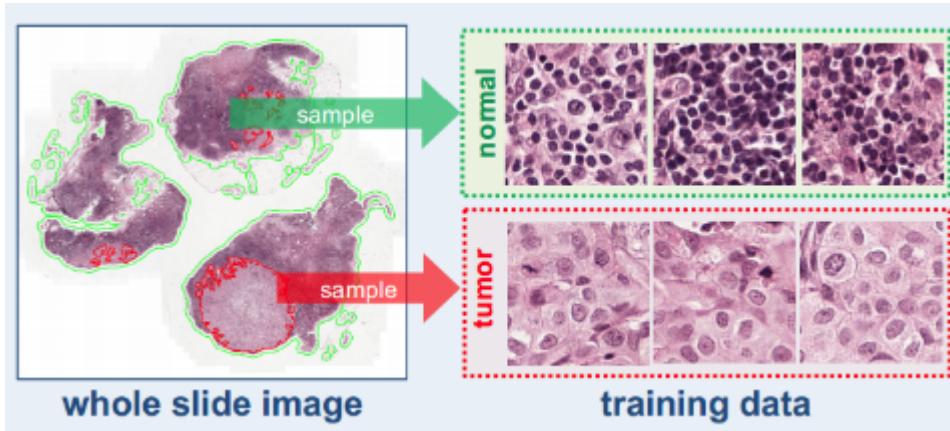


Figure 3.5: Patch extraction from WSI. Tissue and Tumor regions are highlighted with green and red curves respectively. Area within green region but outside of red region is considered as Normal. Extract tumor patches from tumor regions and normal patches from normal regions [1]

Tumor WSIs while test dataset includes 130 unlabeled images. WSIs are very large GigaPixel images, with resolution of  $10^6 \times 10^6$  pixels and raw size of more than 50 GB. Because of large size, it is impossible to load a full image into available computer memory, which makes it infeasible to analyze an entire whole image at once. Due to this limitation, we decided to perform patch based analysis. We randomly extracted thousands of  $256 \times 256$  size patches from ROIs of each WSI image [40], 1k normal & 1k tumor patches from each Tumor slides and 1k normal patches from each Normal slides. In total we extracted 250k patches, 140 normal patches and 110k tumor patches. Figure 3.5 demonstrates the process of extracting patches in detail. Data augmentation techniques have been applied to train model with increasing variety of data and to avoid over fitting of training data. Randomly, we either cropped a  $224 \times 224$  sub-region from original patches or flipped patches horizontally as shown in figure 3.6.

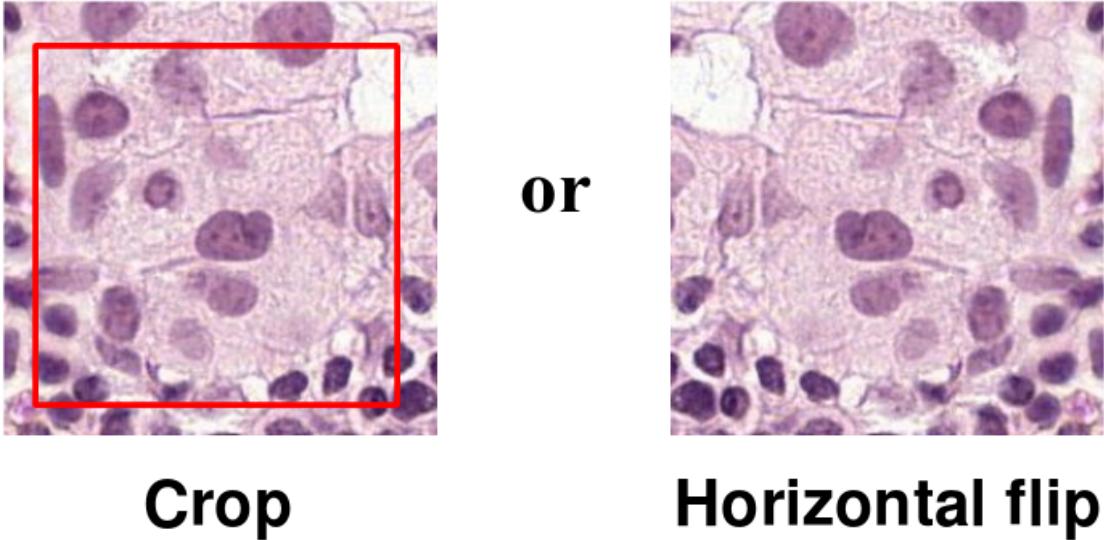


Figure 3.6: Data augmentation techniques

### 3.1.3 Deep ConvNet for tile-based classification

Convolutional neural networks are powerful deep learning based approaches for visual recognition tasks. Starting 2012, major image classification challenges such as ImageNet has been consistently won by deep learning based approaches every year. Recently, deep learning based approaches have also shown state-of-the-art results in medical science challenges such as MICCAI 2013 and ICPR 2012 which shows the potential of applying deep learning to solve real life health care related problems in clinical practices [19]. Classical machine learning based approaches require lots of manual steps for object detection, object segmentation and feature extraction, while deep learning based approach automatically learns high-dimensional complex features just with the use of training data and its labels (e.g. 0 and 1) [1]. In cancer pathology mitosis detection is a very important process as mitosis count is a key indicator of presence and severity of cancer disease. Traditionally, mitosis count

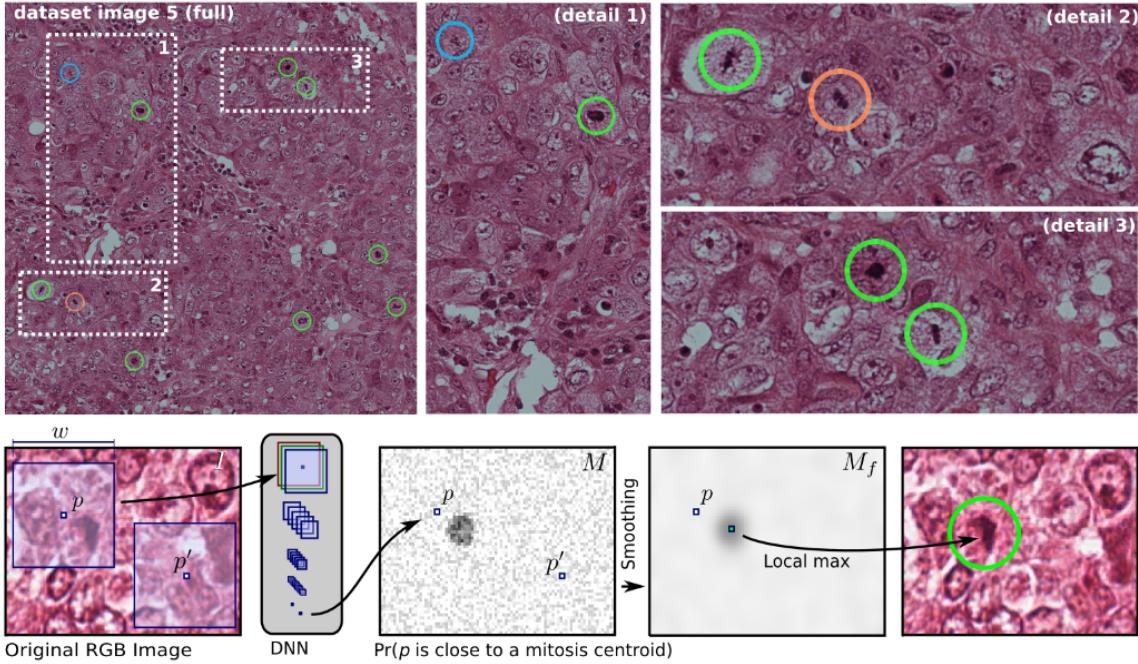


Figure 3.7: *Top*: Images showing detected mitosis. Detected mitosis are circled green (true positives) and red (false positives); cyan denotes mitosis which are not detected. *Bottom*: Mitosis detection framework [19]

has been performed by pathologist but it is a very time consuming process because pathologist has to scan millions of tissues manually. Automating this process has high value in clinical practices because it would be faster, cheaper, more accurate, and more reliable. Classical computing methods developed for mitosis counting are based mainly on pixel classifiers and detecting particular objects from the images. Mitosis detection is quite a challenging process as it is very hard to distinguish between mitotic and non-mitotic nuclei [19]. Thus scientists start developing ConvNet based approach for mitotic detection. Figure 3.7 demonstrates a mitosis detection framework and example of detected mitosis. As stated by [1], ConvNets yield robust hierarchies of features unlike computer vision approaches that rely on hand crafted features from images and videos. [1] also stated that, in ConvNet based approach it is easy to

Model-D1	
<b>Optimizer</b>	<i>SGD with exponential decay (learning rate decay factor = 0.1)</i>
<b>learning rates</b>	<i>Exponential decay: 1*e-02, 1*e-03, 1*e-04</i>
<b>Batch size</b>	<i>32</i>
<b>GPUs</b>	<i>2</i>
<b>Data</b>	<i>250K patches (140K -ve &amp; 110K +ve) with channel wise mean subtraction and pixel range [0.0, 1.0]</i>

Figure 3.8: Model D-1 training details

derive both global and local features as it holds both types of features in the form of non-linear global to local pyramid. It is being observed that there is a huge success for deep learning based methods in categorization of whole images, detecting objects in images or video frames, speech recognition. However, tasks like tumor segmentation in computer aided diagnosis must be practically accurate to each pixel and hence raises the need for robust and highly accurate methods.

This section explains the use of Deep ConvNet for patch based classification. We trained a Deep classification model D-1 from scratch to distinguish between Tumor and Negative patches. Figure 3.8 displays the detailed information about training model D-1. During training, our model uses as input 256 x 256 size patches extracted from positive and negative regions of WSIs as explained in section 3.1.2. We evaluated the performance of many well-known deep learning network architectures for this classification task and then adopted GoogLeNet (Inception-V3) as our deep network structure since it is generally faster and more stable than other networks. The network structure of GoogLeNet consists of 27 layers in total and more than 6 million parameters [10]. Table 3.1 lists detailed information about Inception-V3 network

including types of layer, patch size, and stride and input size of each layer. Also, figure 3.10 depicts layer by layer architecture of Inception-V3 network. For training, we used mini-batch Stochastic Gradient Descent (SGD) as optimizer, Softmax cross entropy as the loss function, and exponential decay mechanism for managing the learning rate. We set initial learning rate to 0.01 and decreased the value by the factor of 10 after each 60k iterations, in order to reduce oscillation and avoid divergence during model training. We also used batch-normalization [17] to achieve faster training and drop-out [18] to avoid over fitting of training data. Using this setup, we trained deep model D-1 with 2 GPU and a batch size of 32 until it converges. Figure 3.9 illustrates the training loss occurred while training Deep model D-1. Environment used for training includes:

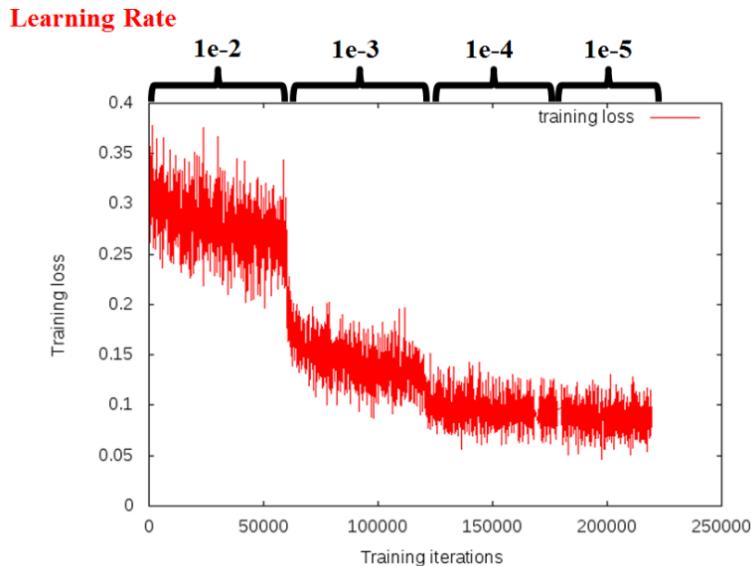


Figure 3.9: Training loss for Deep ConvNet model. Note significant decrease in training loss when learning rate decrease. Here 1 iteration corresponds to 1 batch of 32 images

1. GPU: 2 x 12 GB NVidia K40

2. CPU: 3.4GHz Intel core i7 4770
3. HDD: 7 TB
4. RAM: 16 GB DDR4

<b>type</b>	<b>patch size/stride</b>	<b>input size</b>
conv	3 x 3/ 2	256 x 256 x 3
conv	3 x 3/ 1	128 x 128 x 32
conv padded	3 x 3/ 1	128 x 128 x 32
pool	3 x 3/ 2	128 x 128 x 64
conv	3 x 3/ 1	64 x 64 x 64
conv	3 x 3/ 2	64 x 64 x 80
conv	3 x 3/ 1	32 x 32 x 192
3 x Inception	As in figure 3.11	32 x 32 x 288
5 x Inception	As in figure 3.12	16 x 16 x 768
2 x Inception	As in figure 3.13	8 x 8 x 1280
pool	8x8	8 x 8 x 2048
linear(FC)	logits	1 x 1 x 2048
softmax	classifier	1 x 1 x 2

Table 3.1: Inception-V3 outline [10]

### 3.1.4 Building Tumor probability heat-maps

After training Deep ConvNet model, next step is to generate a tumor probability heat-maps for each WSI. For this, we extracted 256 x 256 size patches from ROIs of each WSI. In total we extracted 7.6 million patches including both train and test WSIs. Then, using trained deep Model D-1, we obtained tumor probabilities for all extracted patches, and embedded predictions of patches from individual WSIs into

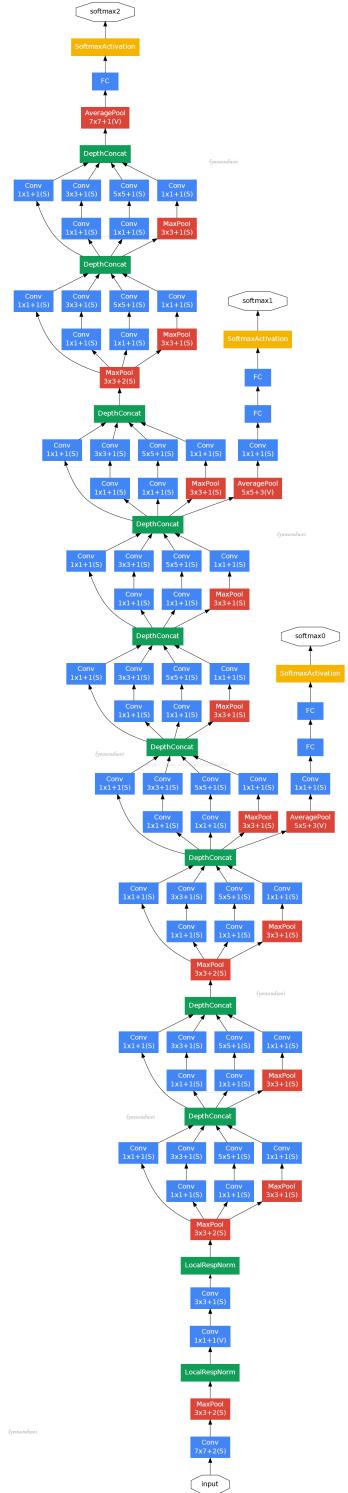


Figure 3.10: Inception-V3 architecture [9]

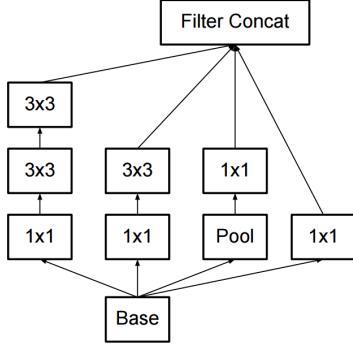


Figure 3.11: Inception modules where each  $5 \times 5$  convolution is replaced by two  $3 \times 3$  convolution [10]

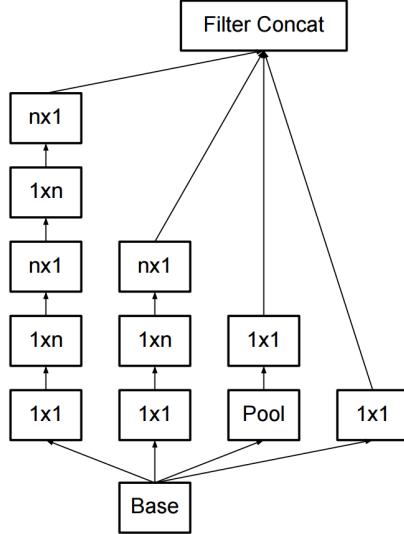


Figure 3.12: Inception modules after the factorization of the  $n \times n$  convolutions. We chose  $n = 7$  for the  $17 \times 17$  grid [10]

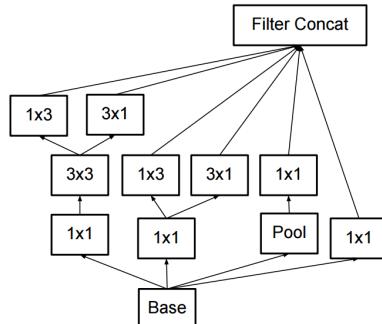


Figure 3.13: Inception modules with expanded the filter bank outputs. This architecture is used on the coarsest ( $8 \times 8$ ) grids to promote high dimensional representations [10]

a single heat-map. In this manner, we built tumor probability heat-maps for every single WSI. In a heat-map, each pixel contains a value between 0 and 1, indicating the probability that the pixel contains a tumor. Figure 3.14 demonstrates the step by step process for building a tumor probability heat-map.

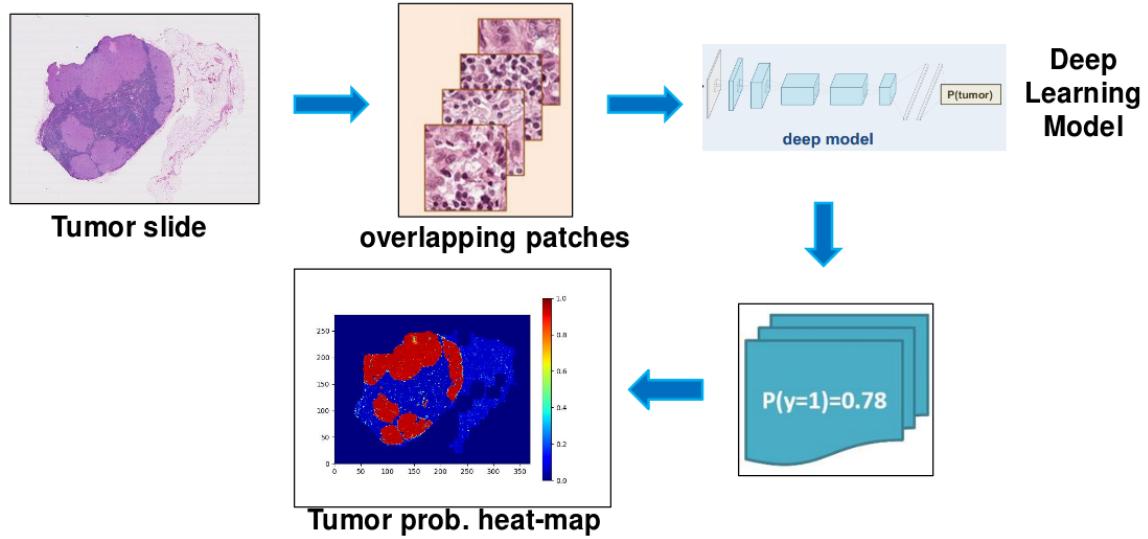


Figure 3.14: Process of building tumor probability heat-map

### 3.1.5 Post-processing on heat-maps for slide-based classification

After building tumor probability heat-maps, we performed post-processing operations on these heat-maps to build the classifier for slide based classification. For the slide-based classification task, the post-processing takes as input a heat-map for each WSI and produces as output a single probability of tumor for the entire WSI. Given a heat-map, we extract 28 geometrical and morphological features from each heat-map, including the percentage of tumor region over the whole tissue region, the area ratio between tumor region and the minimum surrounding convex region, the average prediction values, and the longest axis of the tumor region [1]. We compute

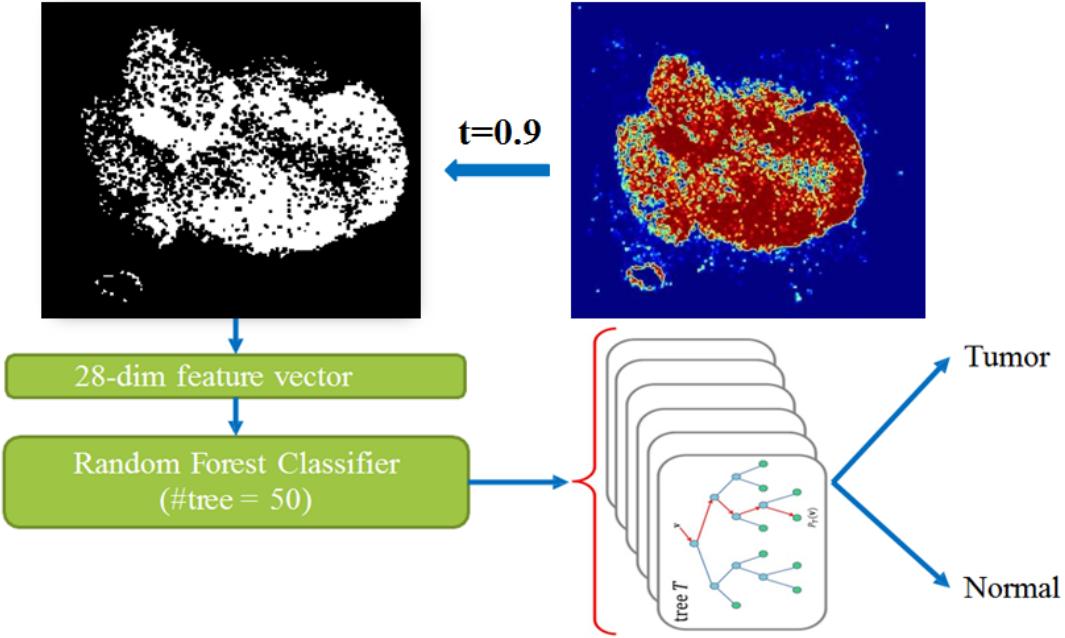


Figure 3.15: Process of extracting features from heat-map

these features over tumor probability heat-maps across all training cases, and we build a Random Forest classifier to discriminate the WSIs with metastases from the negative WSIs. Figure 3.15 depicts step by step process for extracting features from heat-maps. The complete list of features includes [1, 2]:

1. The ratio of tumor region to the tissue region
2. The longest axis in the largest tumor region
3. Total number of pixels with probability greater than 0.90
4. Tumor area: maximum, mean, variance, skewness, and kurtosis
5. Tumor perimeter: maximum, mean, variance, skewness, and kurtosis
6. Eccentricity of ellipse having same second momentum as region: maximum, mean, variance, skewness, and kurtosis

7. Ratio of pixels in the region to pixels in the total bounding box (“extent”):  
maximum, mean, variance, skewness, and kurtosis
8. Solidity: maximum, mean, variance, skewness, and kurtosis

After analyzing the importance of individual features, top 5 important features turns out to be (t is the threshold value):

1. Feature 11: given  $t = 0.9$ , mean area of tumor regions
2. Feature 10: given  $t = 0.5$ , the longest axis in the largest tumor region
3. Feature 09: given  $t = 0.5$ , ratio of pixels in the region to pixels in the total bounding box (“exten”)
4. Feature 05: given  $t = 0.9$ , eccentricity of the ellipse that has the same second moments as the region. (“eccentricit”)
5. Feature 06: given  $t = 0.9$ , ratio of tumor region to the tissue region

Figure 3.16 displays the feature importance graph showing the level of importance for individual features.

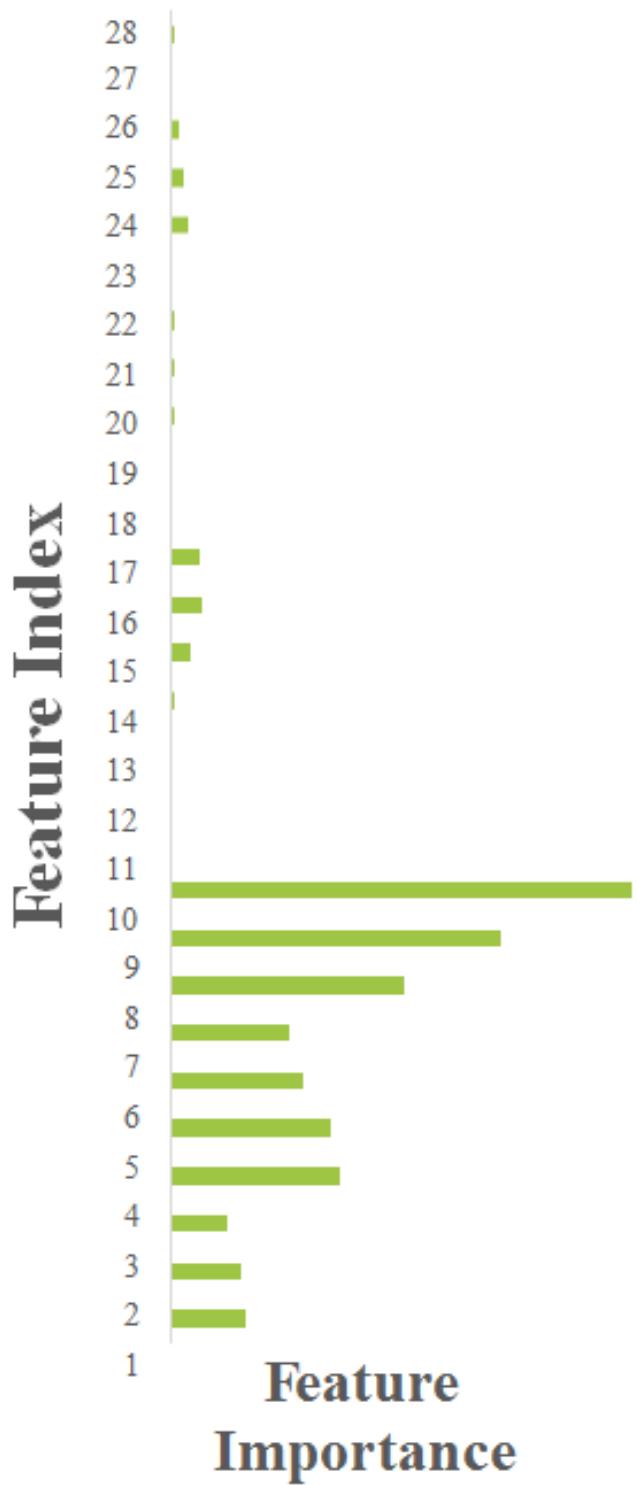


Figure 3.16: Feature importance map

## CHAPTER 4

## EXPERIMENTAL RESULTS

### 4.1 Dataset

For this thesis, we used a breast cancer dataset provided as part of Camelyon'16 [42] grand challenge. It contains a total of 400 whole-slide images (WSIs) of sentinel lymph node from two independent datasets collected in Radboud University Medical Center (Nijmegen, the Netherlands), and the University Medical Center Utrecht (Utrecht, the Netherlands) [42]. The first training dataset consists of 170 WSIs of lymph node (100 Normal and 70 containing metastases) and the second 100 WSIs (including 60 normal slides and 40 slides containing metastases). The ground truth data for the slides containing metastases is provided in two formats: 1. XML files containing vertices of the annotated contours. 2. WSI binary Masks. The test dataset consists of 130 WSIs which are collected from both Universities. Table 4.1 provides specific details about Camelyon'16 dataset. We used Automated Slide Analysis Platform (ASAP), an open source platform, to visualize cancer metastases annotation in whole-slide histopathology images. Figure 4.1 <sup>1</sup> shows an example of visualizing cancer metastases using ASAP.

### 4.2 Experimental setup

To successfully perform experiments on proposed system following requirements were setup at SMILE Lab. The hardware configurations of the system were:

1. CPU:3.4GHz Intel core i7 4770

---

<sup>1</sup><https://camelyon16.grand-challenge.org/data/>

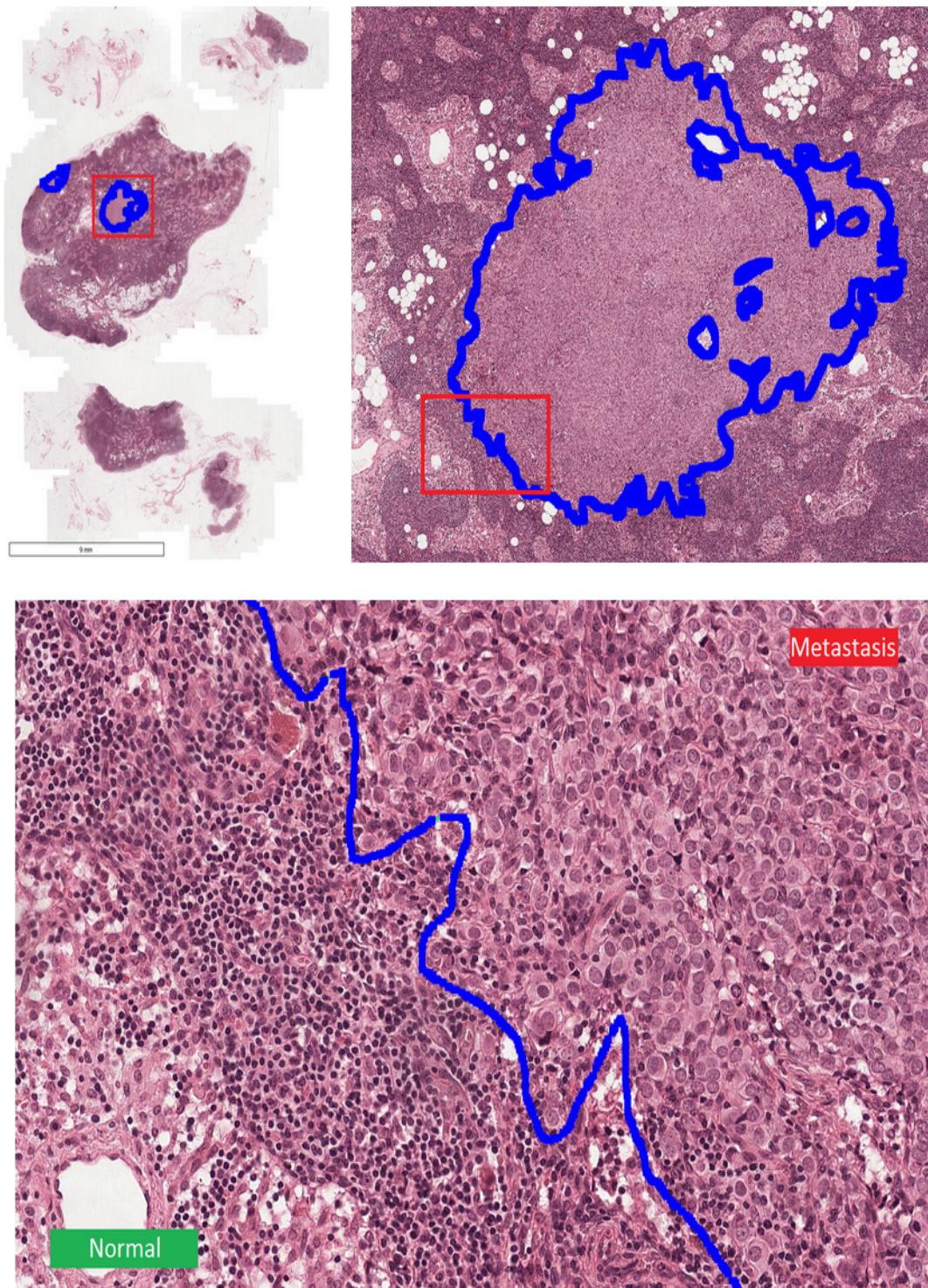


Figure 4.1: Visualizing metastases using ASAP [42]

Source	Train Tumor	Train Normal	Test
Radboud UMC	70	100	80
UMC Utrecht	40	60	50
Total	110	160	130

Table 4.1: Number of slides in the Camelyon'16 dataset

2. RAM: 12GB DDR4
3. 12 GB of NVidia K40 GPU
4. 12 GB of NVidia Ge Force Titan X GPU

The software requirements were:

1. OS: Ubuntu 16.04
2. Programming Languages: Python 3.5
3. Deep Learning libraries : Tensorflow (v0.12.1)
4. Support libraries: OpenSlide, Sci-Kit, NumPy

### 4.3 Evaluation metrics

#### 4.3.1 Receiver Operating Curves (ROC)

The ROC<sup>2</sup> curve helps in the creation of a full and detailed sensitivity vs 1-specificity report. In ROC, the true positive rate (sensitivity) is plotted against the function of the false positive rate (specificity). The sensitivity is the probability of a test outcome being true positive and specificity is the probability of the test outcomes being true negative.

$$\text{Sensitivity or True Positive Rate (TPR)} = \frac{TP}{(TP + FN)} \quad (4.1)$$

$$\text{Specificity or True Negative Rate (TNR)} = \frac{TN}{(TN + FP)} \quad (4.2)$$

---

<sup>2</sup><https://www.medcalc.org/manual/roc-curves.php>

where TP denotes True Postive, TN denotes True Negative,

FP is False Positive, FN is False Negative,

P is Positive and N is Negative

## 4.4 Model D-1 results

### 4.4.1 Training

As explained in section 3.1.3, we trained Deep ConvNet model D-1 to distinguish between tumor and normal patches. We used mini-batch Stochastic Gradient Descent (SGD) as the optimizer, Softmax cross entropy as the loss function and exponential decay mechanism for learning rate. We set the initial learning rate to 0.01 and decreased the value by the factor of 10 after each 60k iterations, in order to reduce oscillation and avoid divergence. In order to eliminate variability in pixel intensity, we adopted standard data normalization techniques. For each training patch, we first brought the pixel intensity values in the range [0.0, 1.0] and then subtracted channel wise mean from each R, G, B channel. The model is trained continually using 2 GPU with batch size 32 until convergence is achieved. After 250k iteration model D-1 obtained a patch based classification accuracy of 97%.

#### 4.4.2 ROC

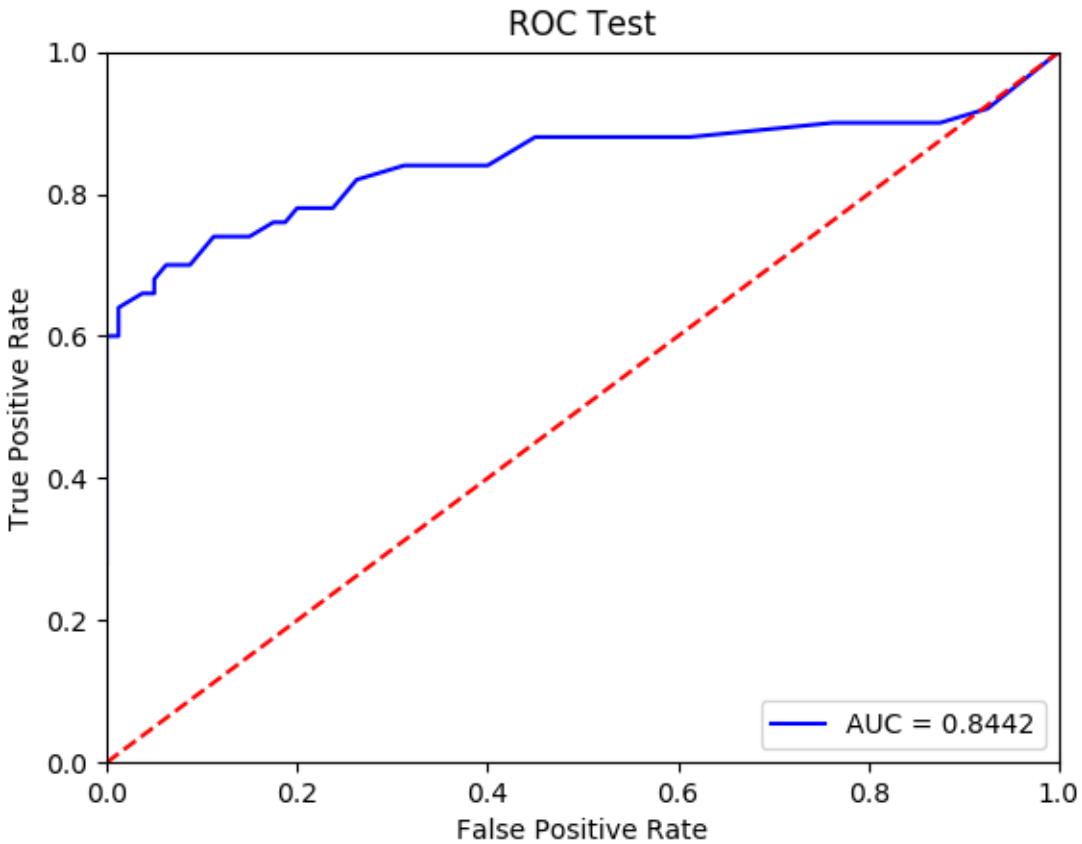


Figure 4.2: ROC curve for Model D-1 predictions on Test dataset

#### 4.5 Ensemble method

As can be seen in Figure 4.2, model D-1 obtained an area under curve (AUC) of 84.42% which is quite low. The low score of model D-1 is because it produces many false positives. Since, heat-maps generated using model D-1 contains many false positives, it is difficult to distinguish Normal WSI heat-maps from Tumor WSI heat-maps. These false positives are due to incomprehensive training data as we have

extracted training patches randomly to reduce number of training samples. Because of this randomness, some difficult negatives patches from the histological mimics of cancer were missed in training data, which results in producing false positives. Figure 4.3 shows an example of false positives in Model D-1 heat-map.

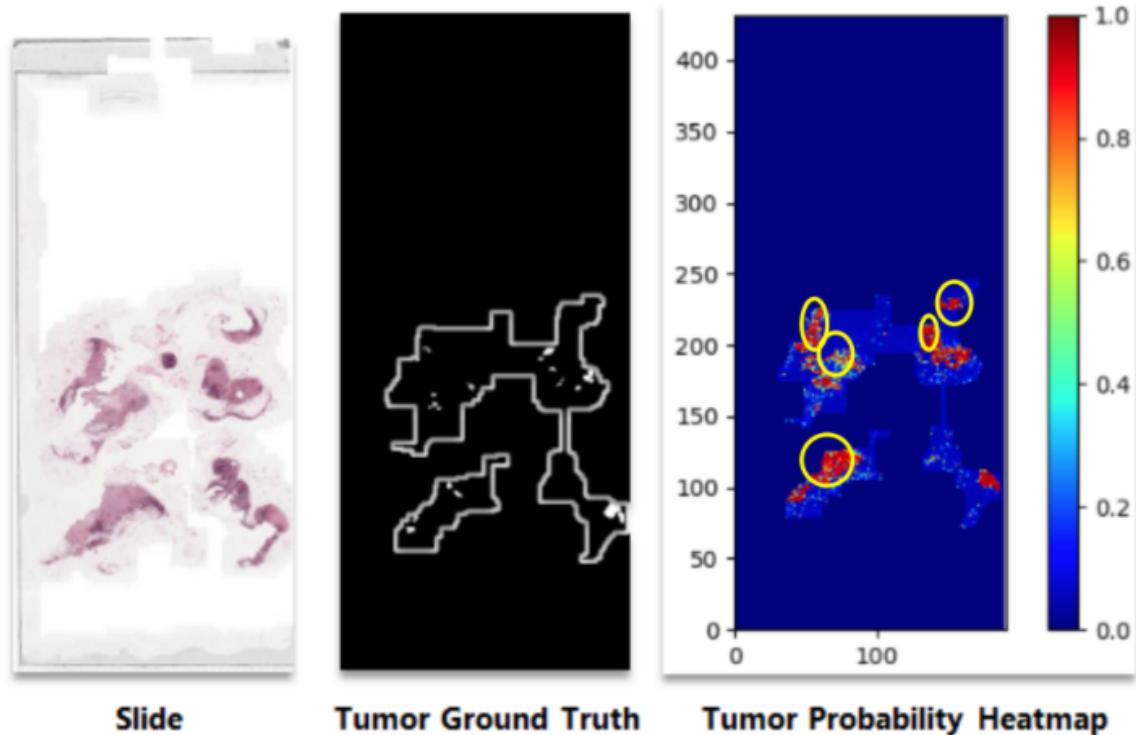


Figure 4.3: Model D-1 false positives. *Left:* Original slide. *Middle:* Ground truth. White pixels inside ROI represents ground truth. *Right:* model D-1 heat-map. Red pixels inside highlighted yellow regions represent false positives

To remove false positives, we extracted appx. 100k additional patches corresponds to these false positive pixels and trained a new model D-2 with this enriched training dataset. The model D-2 is trained using the same environment and training techniques as Model D-1; the only difference being in the dataset. The model D-2 obtained the patch based classification accuracy of 98%. After training model D-2,

we built tumor probability heat-maps for each WSIs with this model. Next, we developed ensemble learning approach using both deep ConvNet models: Model D-1 & Model D-2. In ensemble approach, we built ensemble heat-maps for each WSI by removing false positives in model D-1 heat-maps with model D-2 predictions. Criteria used for preparing ensemble heat-maps is:

$$\text{pr(Ensemble)} = \begin{cases} \text{pr(D-2)}, & \text{pr(D-1)} \geq 0.90 \text{ & } \text{pr(D-2)} < 0.50 \\ \text{pr(D-1)}, & \text{otherwise} \end{cases}$$

Figure 4.4 shows an example of preparing ensemble heat-map for tumor slide Tumor\_074.

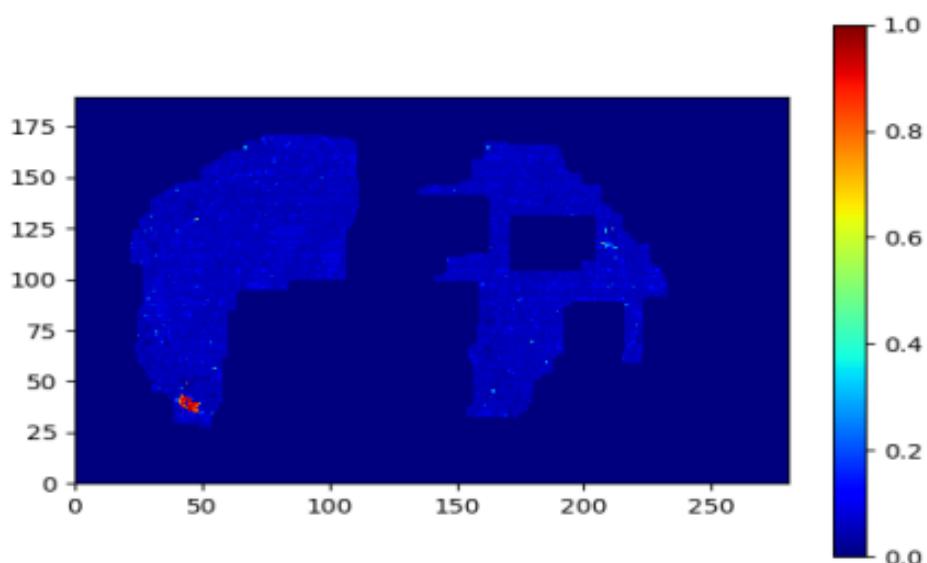
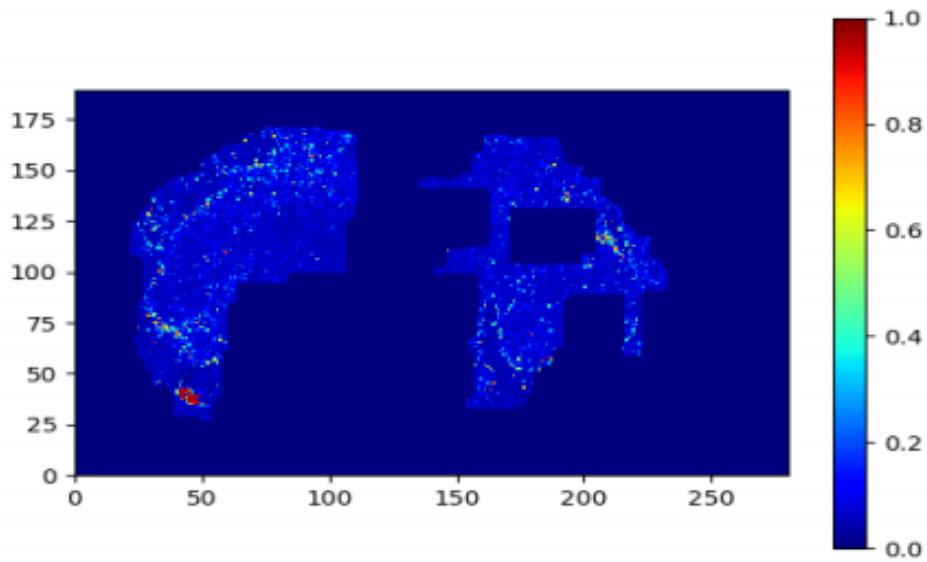


Figure 4.4: Example for preparing ensemble heat-map for slide Tumor\_074. *Top:* Model D-1 heat-map. *Bottom:* Ensemble heat-map

#### 4.5.1 ROC

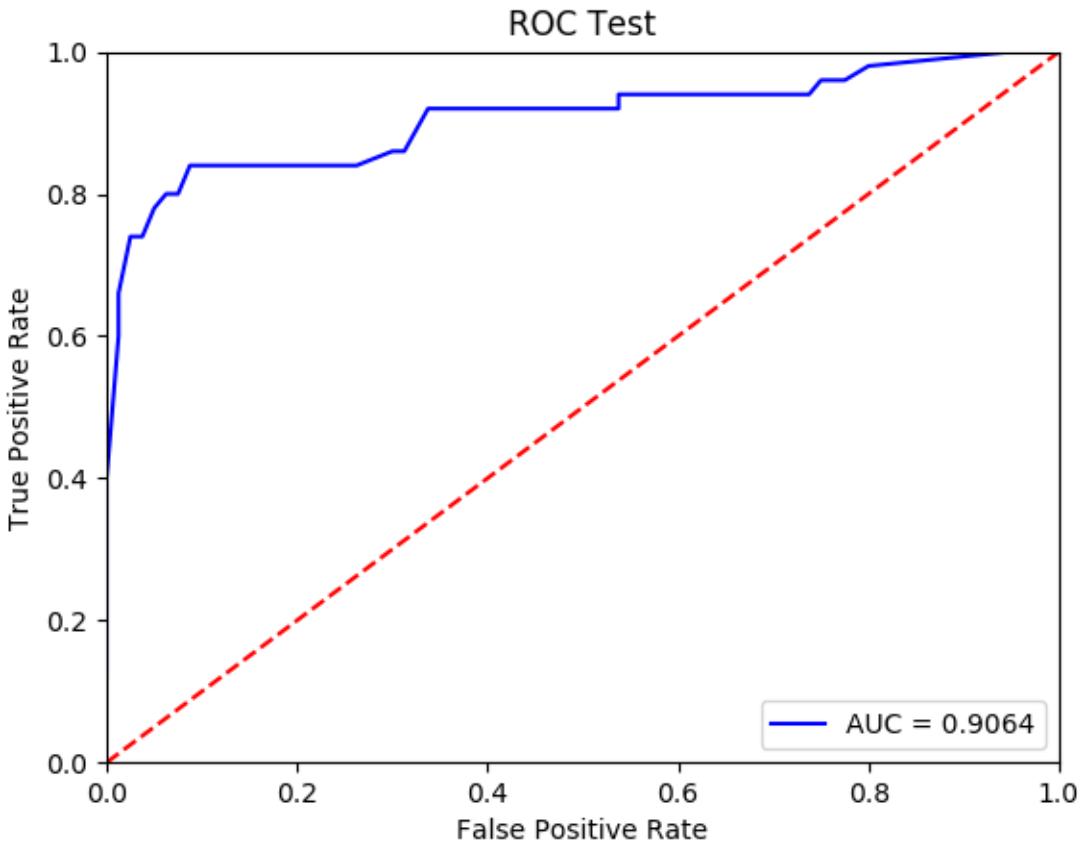


Figure 4.5: ROC curve for ensemble method on Test dataset

### 4.6 Further improvements

#### 4.6.1 Feature selection

As shown in Figure 4.5, using ensemble method, we obtained the area under curve (AUC) of 90.64%. Though, the AUC score obtained by ensemble method is better compared to model D-1, it is still not as good as the winning method of Camelyon'16 grand challenge. In order to further improve the AUC score, we employ

a feature selection method. As discussed in section 3.1.5, we extracted 28 geometrical and morphological features from each heat-map and then trained Random Forest (RF) classifier to distinguish WSIs with metastases from the negative WSIs. After studying feature importance graph shown in Figure 3.16, we found that feature with very low importance does not have any contribution in prediction decisions made by Random Forest classifier. Based on this fact, we removed 4 low importance features from 28 extracted features. In further analysis, using data analysis techniques, we found the features having low correlation with target labels and removed 2 such features. So in total, we removed 6 features from previous 28 features and used remaining 22 features for classification task.

#### 4.6.2 Use of better classifier

Second, in addition to feature selection, we experimented with many different classifiers other than RF: K-Nearest-Neighbour (KNN), Naive Bayes, and Support Vector Machine (SVM). After performing extensive experiments, SVM turns out to be the best amongst all classifiers, hence we have replaced existing RF classifier with SVM, to improve accuracy of feature classification task.

#### 4.6.3 ROC

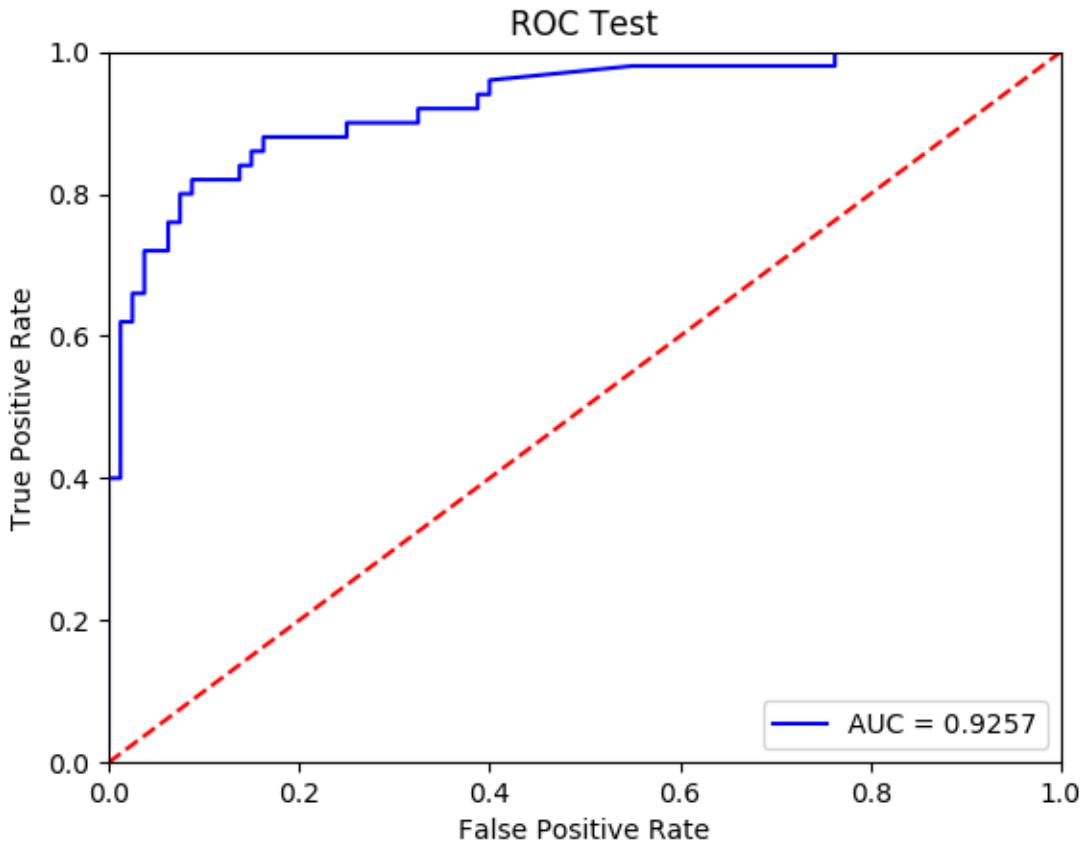


Figure 4.6: Final ROC curve on Test dataset, obtained after applying feature selection and better classifier

#### 4.7 Results comparison

Figure 4.7 compares result obtained by the proposed system with Top-5 methods of the Camelyon'16 grand challenge. The proposed system achieved the final AUC score of 92.57% which beats the winning method of the Camelyon'16 grand challenge, developed to-gather by Harvard & MIT research labs.

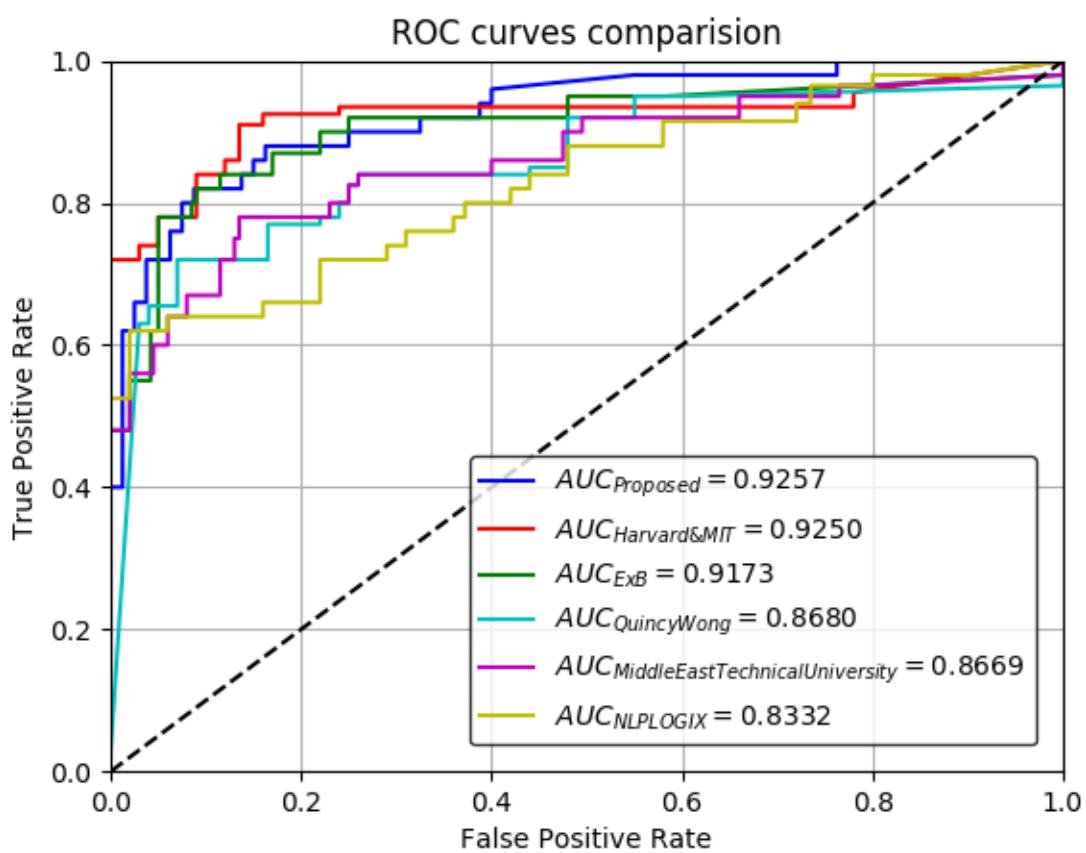


Figure 4.7: ROC comparison with Camelyon16 Top-5 methods

## CHAPTER 5

### CONCLUSION AND FUTURE WORK

The aim of this thesis is to present a deep learning based system for automated detection of metastatic cancer from whole slide images of sentinel lymph nodes. The main challenges of the system we have developed were to enhance the training set to improve previous system by avoiding misclassification of the normal lymph node regions as cancer. Other important key features of the system were to develop state-of-the art deep learning architecture to classify small patches from the large whole slide images, and use of carefully designed post-processing methods for the slide based classification. Classical methods in histopathology were mainly focused on image analysis tasks such as color normalization, nuclear segmentation, and feature extraction [1]. Usually image analysis task alone is not sufficient for cancer classification so in practice after image analysis, machine learning based classification models such as Support Vector Machine and Random Forest are required for end-to-end feature classification task. Since many years, deep learning-based approaches have played a major role in various computer vision competitions, such as ImageNet Large Scale Visual Recognition Competition (ILSVRC). Recently Deep-learning has also emerged as a leading technology in the field of pathology and related research areas in medical science [19]. In contrast to classical machine learning based approaches, deep-learning based approach does not require manual steps for object detection, object segmentation and feature extraction as it automatically learns high-dimensional complex features, just with the use of training data and its labels (e.g. 0 and 1) [1]. The proposed method utilizes GoogLeNet, a 27-layer deep network architecture, and obtained near human-

level classification performance on the breast cancer test data. As tested by [23], the errors made by deep learning based systems are not strongly correlated with the errors made by human pathologist. Thus, although the pathologist alone is currently superior to deep learning system alone, combining deep learning with the pathologist produced a major reduction in pathologist error rate, reducing it from over 3 percent to less than 1 percent. Based on our results, we can conclude that integrating deep-learning based approaches into clinical practices can bring vast improvements in speed, accuracy, reproducibility, reliability and clinical value of pathological diagnoses.

Based on our evaluations, a future research direction could be to train and evaluate the performance of our system on another large scale open source cancer data such as TCGAs lung cancer dataset. Another important step in future research would be to integrate staining normalization process into proposed classification pipeline. Stain normalization is a crucial part of building a generalize deep learning model for identifying metastatic regions as it eliminates the stain variability in WSIs, induced by different staining techniques. As shown by [2], stain normalization process has a great potential to significantly improve classification performance of a deep model, which indicates that with stain normalization on board, proposed system could produce even better results and could be extended to any breast cancer digitized images without institutional re-calibration.

## REFERENCES

- [1] D. Wang, A. Khosla, R. Gargeya, H. Irshad, and A. H. Beck. “Deep Learning for Identifying Metastatic Breast Cancer”, arXiv preprint arXiv:1606.05718, 2016.  
Copy at <http://j.mp/2o6FejM>
- [2] Chen, Richard et al.”Identifying Metastases in Sentinel Lymph Nodes with Deep Convolutional Neural Networks. CoRR abs/1608.01658 (2016): n. pag.
- [3] Ruoyu Li, Junzhou Huang. “Fast Regions-of-Interest Detection in Whole Slide Histopathology Images”, 1st International Workshop on Patch-based Techniques in Medical Imaging, PMI15, Munich, Germany, October 2015.
- [4] R. S. Cotran, V. Kumar, T. Collins, and S. L. Robbins. “Robbins pathologic basis of disease”, 1999.
- [5] B. J. Czerniecki, A. M. Scheff, L. S. Callans, F. R. Spitz, I. Bedrosian, E. F. Conant, S. G. Orel, J. Berlin, C. Helsabeck, D. L. Fraker, et al. “Immunohistochemistry with pancytokeratins improves the sensitivity of sentinel lymph node biopsy in patients with breast carcinoma”, Cancer, 85(5):10981103, 1999.
- [6] Jiawen Yao, Dheeraj Ganti, Xin Luo, Guanghua Xiao, Yang Xie, Shirley Yan and Junzhou Huang. “Computer-assisted Diagnosis of Lung Cancer Using Quantitative Topology Feature”, 6th International Workshop on Machine Learning in Medical Imaging, MLMI’15, Munich, Germany, October 2015.
- [7] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. “Gradient-based learning applied to document recognition.” Proceedings of the IEEE, November 1998
- [8] A. Krizhevsky, I. Sutskever, and G. E. Hinton. “Imagenet classification with deep convolutional neural networks”, In NIPS, 2012.

- [9] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. “Going deeper with convolutions”, in CVPR, pp. 1-9, 2015.
- [10] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. “Rethinking the inception architecture for computer vision”, arXiv preprint arXiv:1512.00567, 2015.
- [11] K. Simonyan and A. Zisserman. “Very deep convolutional networks for large-scale image recognition”, CoRR, vol. abs/1409.1556, 2014.
- [12] He K, Zhang X, Ren S, Sun J. “Deep Residual Learning for Image Recognition”, CVPR 2015.
- [13] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al. “Imagenet large scale visual recognition challenge”, International Journal of Computer Vision, 115(3):211252, 2015.
- [14] N. Otsu. “A Threshold Selection Method from Gray-level Histograms”, IEEE Transactions on Systems, Man and Cybernetics, 9(1):6266, 1979.
- [15] P.J. Tadrous. “Digital stain separation for histological images”, Journal of Microscopy, Vol. 240, 2010, pp. 164 - 172.
- [16] Huey Nee Lim, Mohd Yusoff Mashor, Nadiatun Zawiyah Supardi, Rosline Hassan. “Color and Morphological Based Techniques on White Blood Cells Segmentation”, in Proc. IEEE 2nd Intl. Conf. on Biomed. Engineering, 2015.
- [17] S. Ioffe and C. Szegedy. “Batch normalization: Accelerating deep network training by reducing internal covariate shift”, CoRR, vol. abs/1502.03167, 2015.
- [18] Nitish Srivastava , Geoffrey Hinton , Alex Krizhevsky , Ilya Sutskever , Ruslan Salakhutdinov. “Dropout: a simple way to prevent neural networks from overfitting”, The Journal of Machine Learning Research, v.15 n.1, p.1929-1958, January 2014

- [19] D. C. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber. “Mitosis detection in breast cancer histology images with deep neural networks”, In Medical Image Computing and Computer-Assisted InterventionMICCAI 2013, pages 411418. Springer, 2013.
- [20] F. Ghaznavi, A. Evans, A. Madabhushi, and M. Feldman. “Digital imaging in pathology: whole-slide imaging and beyond”, Annual Review of Pathology: Mechanisms of Disease, 8:331359, 2013.
- [21] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener. “Histopathological image analysis: a review. Biomedical Engineering”, IEEE Reviews in, 2:147171, 2009.
- [22] H. Irshad, A. Veillard, L. Roux, and D. Racoceanu. “Methods for nuclei detection, segmentation, and classification in digital histopathology: A reviewcurrent status and future potential”, Biomedical Engineering, IEEE Reviews in, 7:97114, 2014.
- [23] D. L. Weaver, D. N. Krag, E. A. Manna, T. Ashikaga, S. P. Harlow, and K. D. Bauer. “Comparison of pathologist detected and automated computer-assisted image analysis detected sentinel lymph node micro metastases in breast cancer”, Modern pathology, 16(11):11591163, 2003.
- [24] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard and L. D. Jackel. “Handwritten digit recognition with a back-propagation network”, in Touretzky, David (Eds), Advances in Neural Information Processing Systems (NIPS 1989), 2, Morgan Kaufman, Denver, CO, Video, 1990
- [25] C. Callau et al. “Evaluation of cytokeratin-19 in breast cancer tissue samples: a comparison of automatic and manual evaluations of scanned tissue microarray cylinders”, Biomed. Eng. OnLine, vol. 14. No. 2, p. S2, 2014.

- [26] Abdel-Hamid, O., Mohamed, A., Jiang, H., and G. Penn. “Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition”, Proc.ICASSP, 2012.
- [27] S. Jaffer and I. J. Bleiweiss. “Evolution of sentinel lymph node biopsy in breast cancer, in and out of vogue?”, Advances in anatomic pathology, 21(6):433442, 2014.
- [28] Zheng Xu, Junzhou Huang. “Detecting 10,000 Cells in One Second”, In Proc. of the 19th Annual International Conference on Medical Image Computing and Computer Assisted Intervention, MICCAI’16, Athens, Greece, October 2016.
- [29] Sheng Wang, Jiawen Yao, Zheng Xu, Junzhou Huang. “Subtype Cell Detection with an Accelerated Deep Convolution Neural Network”, In Proc. of the 19th Annual International Conference on Medical Image Computing and Computer Assisted Intervention, MICCAI’16, Athens, Greece, October 2016.
- [30] Jiawen Yao, Sheng Wang, Xinliang Zhu, Junzhou Huang. “Clinical Imaging Biomarker Discovery for Survival Prediction on Lung Cancer Imaging Genetic Data”, In Proc. of the 19th Annual International Conference on Medical Image Computing and Computer Assisted Intervention, MICCAI’16, Athens, Greece, October 2016.
- [31] Zheng Xu, Junzhou Huang. “Efficient Lung Cancer Cell Detection with Deep Convolution Neural Network”, 1st International Workshop on Patch-based Techniques in Medical Imaging, PMI’15, Munich, Germany, October 2015.
- [32] Xinliang Zhu, Jiawen Yao, Feiyun Zhu and Junzhou Huang. “WSISA: Making Survival Prediction from Whole Slide Pathology Images”, CVPR 2017.
- [33] R. C. Gonzalez and R. E. Woods. “Digital Image Processing”, Upper Saddle River, NJ, USA: Pearson, 2008.

- [34] Zhu, Feiyun and Wang, Ying and Xiang, Shiming and Fan, Bin and Pan, Chunhong. “Structured sparse method for hyperspectral unmixing”, ISPRS Journal of Photogrammetry and Remote Sensing, 88:101-118, 2014.
- [35] Zhu, Feiyun and Liao, Peng and Zhu, Xinliang and Yao, Yaowen and Huang, Junzhou. “Cohesion-based Online Actor-Critic Reinforcement Learning for mHealth Intervention”, arXiv:1703.10039, 2017.
- [36] Li, Haichang and Wang, Ying and Xiang, Shiming and Duan, Jiangyong and Zhu, Feiyun and Pan, Chunhong. “A label propagation method using spatial-spectral consistency for hyperspectral image classification”, International Journal of Remote Sensing, 37(1):191-211, 2016.
- [37] Cheng, Guangliang and Wang, Ying and Zhu, Feiyun and Pan, Chunhong. “Road extraction via adaptive graph cuts with multiple features”, Image Processing (ICIP), IEEE International Conference on, 3962-3966, 2015.
- [38] G. H. Lyman, A. E. Giuliano, M. R. Somerfield, A. B. Benson, D. C. Bodurka, H. J. Burstein, A. J. Cochran, H. S. Cody, S. B. Edge, S. Galper, et al. “American society of clinical oncology guideline recommendations for sentinel lymph node biopsy in early-stage breast cancer”, Journal of Clinical Oncology, 23(30):77037720, 2005.
- [39] G. H. Lyman, S. Temin, S. B. Edge, L. A. Newman, R. R. Turner, D. L. Weaver, A. B. Benson, L. D. Bosserman, H. J. Burstein, H. Cody, et al. “Sentinel lymph node biopsy for patients with early-stage breast cancer: American society of clinical oncology clinical practice guideline update”, Journal of Clinical Oncology, 32(13):13651383, 2014.
- [40] Feiyun Zhu and Bin Fan and Xinliang Zhu and Ying Wang and Shiming Xiang and Chunhong Pan. “10,000+ Times Accelerated Robust Subset Selection ARSS”, Proc. Assoc. Adv. Artif. Intell. (AAAI), 3217-3224, 2015.

- [41] Digital pathology, In Wikipedia. Retrieved April 25, 2017, from  
[https://en.wikipedia.org/wiki/Digital\\_pathology](https://en.wikipedia.org/wiki/Digital_pathology)
- [42] Camelyon 2016 grand challenge. Retrieved April 25, 2017, from  
<https://camelyon16.grand-challenge.org/>
- [43] Cheng, Guangliang and Zhu, Feiyun and Xiang, Shiming and Pan, Chunhong. “Road Centerline Extraction via Semisupervised Segmentation and Multidirection Nonmaximum Suppression”, IEEE Geoscience and Remote Sensing Letters, 13(4):545-549, 2016.
- [44] Cheng, Guangliang and Wang, Ying and Gong, Yongchao and Zhu, Feiyun and Pan, Chunhong. “Urban road extraction via graph cuts based probability propagation”, 2014 IEEE International Conference on Image Processing (ICIP), 5072-5076, 2014.
- [45] Cheng, Guangliang and Zhu, Feiyun and Xiang, Shiming and Wang, Ying and Pan, Chunhong. “Accurate urban road centerline extraction from VHR imagery via multiscale segmentation and tensor voting”, Neurocomputing, 205:407-420, 2016.
- [46] U.S. Breast Cancer Statistics. Retrieved April 28, 2017, from  
[http://www.breastcancer.org/symptoms/understand\\_bc/statistics](http://www.breastcancer.org/symptoms/understand_bc/statistics).

## BIOGRAPHICAL STATEMENT

Arjun Punabhai Vekariya received his Bachelors of Technology in computer Science in 2012 from Nirma University, Ahmedabad, India. After bachelors, he worked as a Software Engineer for 3.5 years at Samsung Research Institute, Noida, India. During his tenure with Samsung, he received multiple opportunities to work overseas at Samsung HQ in South Korea and Samsung Vietnam Mobile Center (SVMC) in Hanoi, Vietnam. He started his Masters in Computer Science at The University of Texas at Arlington in Fall 2015 and joined Dr.Junzhou Huangs research lab in Summer of 2016. His areas of interest includes Deep Learning, Machine Learning, Computer Vision, Data Science & Analytics and Android application development.