CATANET: A RESEARCH-BASED PRODUCT FOR METASTATIC BREAST CANCER STAGE PREDICTION USING DEEP LEARNING

by

DR. MUHAMMAD SHAHZAD, USAMA BAIG,
MUHAMMAD MAHAD KHAN, KHALID .

Presented to the Faculty of the School of Electrical Engineering and Computer Science at National University of Sciences and Technology in Partial Fulfillment of the Requirements

for the Degree of

BACHELOR OF SOFTWARE ENGINEERING

NATIONAL UNIVERISTY OF SCIENCES AND TECHNOLOGY, PAKISTAN

May 2019

DEDICATION

To Allah Almighty

&

To our Parents & Faculty

ACKNOWLEDGEMENTS

We are deeply thankful to our Advisor and Co-Advisor, Dr. Mohammad Shahzad and Dr. Qaiser Riaz for helping us throughout the course in accomplishing our final project. Their guidance, support and motivation enabled us in achieving the objectives of the project.

Table of Contents

List of Figures

List of Tables

ABSTRACT

Pathologists spend hours to detect metastasis in hematoxylin and eosin (H&E) stained whole-slide images of lymph node sections. Digital pathology is a new, rapidly expanding field of medical imaging. In digital pathology, whole-slide scanners are used to digitize glass slides containing tissue specimens at high resolution (up to 160nm per pixel).

'CataNet' focuses on the detection of micro and macro metastasis in lymph node digitized images. This subject is highly relevant; lymph node metastases occur in most cancer types (e.g. breast, prostate, and colon). The lymph nodes in the underarm are the first place breast cancer is likely to spread. The project is divided into two phases.

For Slide-Level Classification we detect and localize the cancer cells in whole slide images. For this purpose, we first extract Region of Interest with Image processing, then construct Training Data tiles from ROI, Train Deep Convolutional model for tile-based classification, building tumor probability heat-maps using trained model and report detected tumor. At the end of phase 1, given a hematoxylin and eosin (H&E) stained whole-slide images of lymph node, we are able to tell whether the image contains tumor or not, and if present then localize it within the image.

For Patient-Level Classification we determine the pathologic N-stage (pN-stage) label per patient by extracting the features from the probability maps generated by Phase 1, training a classifier on these features to determine slide level metastasis label and the determining the pN stage of patient using predetermined criteria. At the end of phase 2, given at least 5 whole-slide images of a patient, the product shall be able to determine the pN-stage as provided by the Union for International Cancer Control (UICC).

The aim is to contribute our minor share for the service of this country by building this tool indigenously and helping the local research community.

Chapter 1

INTRODUCTION

1.1 CANCER

Cancer is a non-communicable group of diseases that involves the abnormal growth of cells in such a way that it takes over normal cells and that the body cannot control it. It becomes difficult for the human body to function properly when this happens. Cancer can be cured very effectively; 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. Malignant cells can be found in the colon, the breast, the lungs, and also in the blood of a patient. Generally, all cancers are similar. They usually vary in the way they grow and spread. To explain cancer, we must understand what cancer cells are and how they work. The cell is the elementary structural, functional and biological unit of all living organisms. Each cell performs a specific task. Cells are damaged in some situations, or it may get worn out. In this case, the cell is destroyed and is substituted with a new healthy cell. One of the most important characteristics of a cell is the natural way 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. Cells become cancerous 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 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 quickly, whereas others grow slowly. Different cancers respond to treatment in different ways. 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 its origin. This is known as the cancer stage. The farther the cancer has travelled from its origin, the higher the stage the cancer of the patient is, which could be either in stage 1,2,3 or 4. This information helps the doctor decide which treatment to choose. Cancer at 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 cancer is a cancer that grows from breast tissue. It is said to be the second leading cause of death among women. It is also assessed that over 40,000 women diagnosed with breast cancer die every year in the United States. Breast cancer originates in the cells of the lobules, that are the milk producing glands, 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 of cancer.

1.2.1 Types of Breast Cancer

1.2.1.1 **Ductal carcinoma in situ:**

Ductal Carcinoma in Situ (DCIS) is a non-invasive breast cancer where abnormal cells reside in the lining of the breast milk duct. Carcinoma in Situ is helps us in describing 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 neighboring breast tissue. In early stages, this type of cancer is highly treatable, however, if left untreated or undetected, it can spread into the neighboring breast tissue. This type of breast cancer is classified under stage 0.

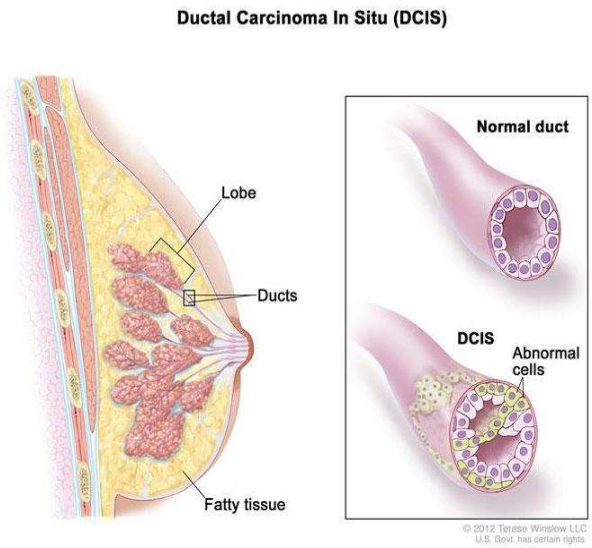

Figure 1: Ductal Carcinoma In Situ

1.2.1.2 **Invasive ductal carcinoma:**

Invasive Ductal Carcinoma means that damaged cells that were created in the lining of the breast milk duct have spread beyond the ducts and invaded neighboring tissue. It is the most common type of

breast cancer. Around 70-80% patients are diagnosed of this type of breast cancer, also this type of cancer is most common in men.

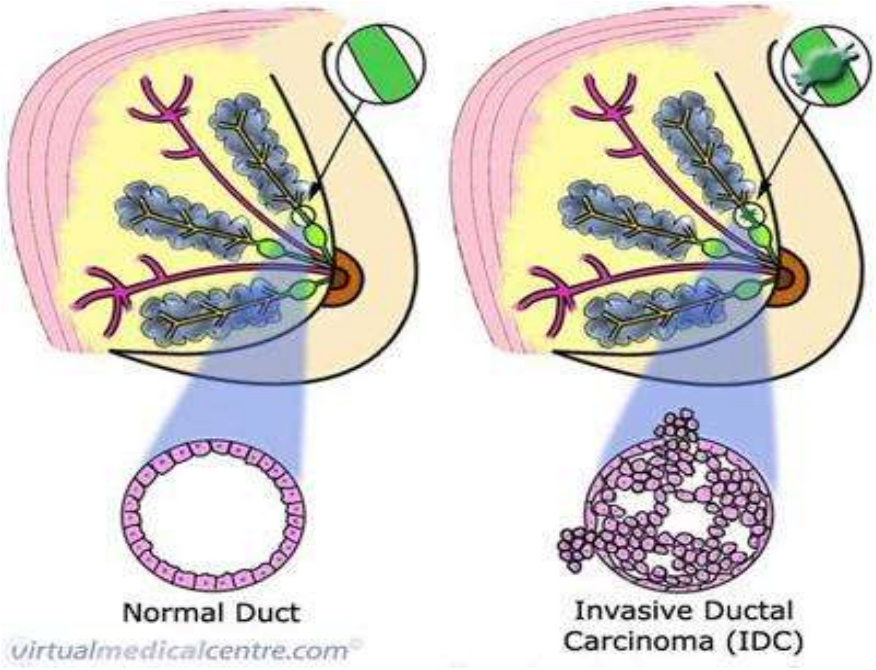

Figure 2: Invasive Ductal Carcinoma

1.2.1.3 **Triple negative breast cancer:**

Triple negative breast cancer means that the cancerous cells are negative for progesterone, estrogen, and HER2/neu receptors, which are receptors known to fuel most breast cancer growth. In this type of cancer, common treatments such as hormone therapy and drugs that hit these receptors become ineffective. In this case chemotherapy is an effective way to treat this kind of breast cancer. Also, 10-20% of breast cancer is known to be triple negative. Usually younger people are affected by this type of cancer or patients with a BRCAi gene mutation.

1.2.1.4 **Inflammatory breast cancer:**

Inflammatory breast cancer is not very common form of breast cancer. This type of cancer does not produce any distinct tumor or hump and is isolated within the breast. It is an aggressive and rapidly 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 cancer is classified as stage 3 breast cancer and requires aggressive treatment.

1.2.1.5 **Metastatic breast cancer:**

Metastatic breast cancer is cancer that spreads beyond the breast, into other parts of our 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.

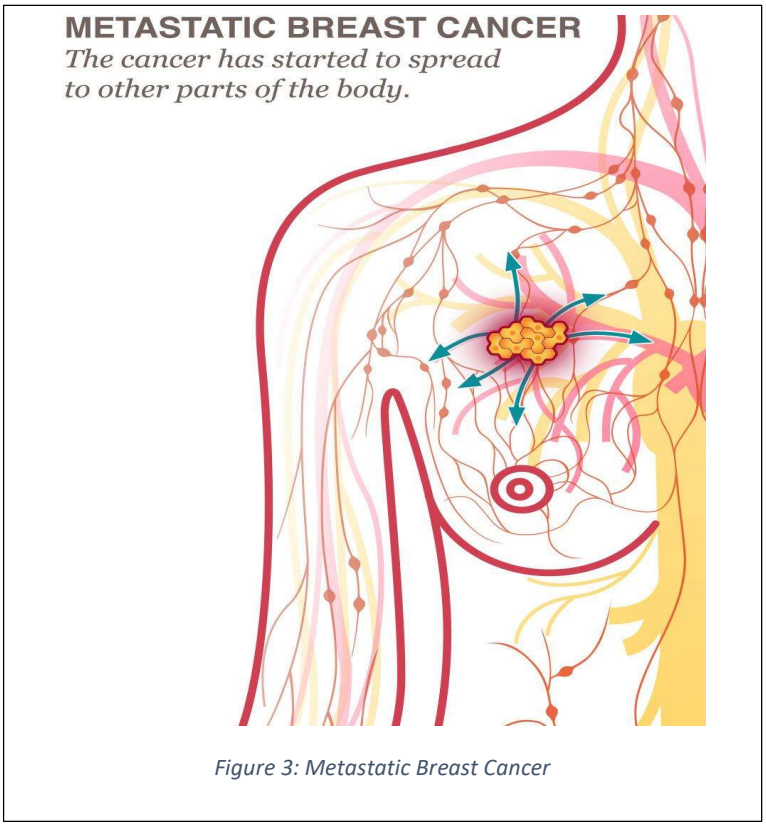

Figure 3: Metastatic Breast Cancer

1.2.2 Diagnosis

1.2.2.1 **Mammogram:**

A mammogram is an x-ray of the breast. This lets a qualified specialist to examine the breast tissue for any symptoms 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 of 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.

1.2.2.2 **Ultrasound:**

When abnormalities in patient's screening mammogram are found, a breast ultrasound is recommended. 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 calculate a digital image of what is going on inside the breast tissue.

1.2.2.3 **MRI:**

MRI is a radio imaging technique that is used to get an understanding of patient's breast internal structure. 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.

1.2.2.4 **Biopsy:**

A breast biopsy is the only diagnostic procedure that can define if the suspected area is cancerous. It is a test that removes tissue or sometimes fluid from the suspicious area. The removed cells are studied under a microscope and further tested to check for the existence of breast cancer. Biopsies are of three types:

1. Fine needle aspiration
2. Core-needle
3. 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 necessary for the patient to have a positive relationship with the doctor. Patient must also 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 someday 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, or the platelets. It is also classified based on its rate of growth, which is either acute for fast growing cancer, or 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 AND CHALLENGES IN DIAGNOSIS

Pathology is the medical field that mainly 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. The recent advents in computing pathology also made it possible to predict survival of a cancer patient. 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, if we look behind at the past, the microscope was the primary tool used by pathologists, ever since mid-19th century. The images formed by these microscopes had many limitations after qualitative visual analysis of them, which included no 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. 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. To improve accuracy of metastatic detection, many clinical laboratories have tried applying proteins such as immune-histochemistry for pancytokeratins 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.

1.3.1 Digital Pathology

Pathology is a medical field that is over 150 years old, which has progressed by leaps and bounds thanks to the advent of digital pathology. 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. 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. 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 supersedes or is on 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 led to a longer adoption curve than compared to other medical specialties that have gone totally digital, like radiology. 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. 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. In machine learning, neural networks are made of large number of layers (deep layers) to solve problems related 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 revolutionized AI industry and swiftly neural network became the hot favorite topic in the field of machine learning research. Many problems in different fields of Computer Science have been solved with the help of various Deep Learning architectures. Deep Learning architectures have been applied to problems in fields like computer vision, automatic speech recognition, 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 main feature 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, which has local connections and shared weights. It is one of the most significant construct of machine learning when it comes to the modern problems, and it has been very popularly used to solve image recognition tasks, in the field of Computer Vision.

The treatment and management of breast cancer is determined by the disease stage. A central component of breast cancer staging involves the microscopic examination of lymph nodes adjacent to the breast for evidence that the cancer has spread, or metastasized. This process requires highly skilled pathologists and is fairly time-consuming and error-prone, particularly for lymph nodes with either no or small tumors. Computer assisted detection of lymph node metastasis could increase the sensitivity, speed, and consistency of metastasis detection.

Chapter 2

LITERATURE REVIEW

Pathologists are tasked with providing conclusive disease diagnoses to guide patient treatment and management decisions. Standardized and accurate pathological diagnoses are essential for advancing precision medicine. Since the 19th century, pathologists used primary tool like microscope to make diagnoses. Microscopic images have many qualitative limitations like; lack of standardization, prone to errors and significant cognitive load required to manually evaluate millions of cells across numerous slides in a normal routine day. Therefore, in recent decades, a lot of work has been done to develop such computational methods that can assist pathologists in the analysis of microscopic images.

In past few years, deep convolutional neural networks (ConvNets) have shown significant improvements on a variety of different computer vision tasks on natural images, e.g. image classification, object detection and semantic segmentation. Similarly, other favorable studies have also applied deep ConvNets to analyze medical images and Whole Slide Images in particular [1, 2, 3, 6, 4, 5], among which [1] won the CAMELYON'16 challenge [1] for metastasis detection. Due to the extremely large size of Whole Slide Images, most of the researchers extracted small patches from WSIs, and then trained their models to classify the patches into normal or cancerous regions. A probability map for the given WSI being tumor or cancerous at patch level was obtained later which enabled to detect the metastasis. However, small neighboring patches often share spatial correlations. But the patches were extracted and trained independently and the spatial correlations were not considered during training. For this reason, for this reason, during inference time, the predictions over neighboring patches were inconsistent. But later Yi

Li and et al [7] introduced NCRF (conditional random field) layer at the end of their model to incorporate for the spatial correlation. They trained whole deep ConvNet end-to-end with standard back propagation algorithm with very small computational overhead from CRF layer. CRF layer also helped them in ConvNet's feature extraction. They showed better results as compared to the baseline method that do not consider spatial correlations. NCRF framework obtained better patch predictions with better visual quality of probability maps.

Chapter 3

Problem Definition

3.1 FACTS & FIGURES

Breast cancer is said to be the second leading cause of death among women. The Age Standardized Mortality Rate (ASMR) for Pakistan (GLOBOCAN 2012 estimates) was 25.2/100,000, highest among South Asian countries. The major reason being delayed clinical evaluation, diagnosis and staging, and absence of timely access to optimum treatment. The early diagnosis of cancer is a vital part of pathology and demands high accuracy. To manually conduct a pathological review, is very time consuming and laborious. But thanks to advancement in Whole Slide Imaging (WSI), the field of digital pathology has exploded. And several, computer based techniques are available for better, faster and cheaper diagnosis. However, sadly in Pakistan, the field has very little application to our knowledge. We would like to associate this backwardness due to 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".

3.2 CONSTRAINTS

Our data set contains Whole Slide Images that are very large in size i.e 1 Giga-pixel images (10^6x10^6). Panel of pathologists have to scan whole image manually which takes up to 30+ hours. This process is prone to error and time consuming. Rapid, cheaper and precise diagnosis methods are required, but classical methods are not robust, they use low level image analysis tasks like:

- Color normalization
- Nuclear segmentation
- Feature extraction

To process whole slide images, resources were required in abundance. Moreover, only 8 false positives are allowed on average upon 10^5 predictions per slide.

Our solution should also be useable by pathologists and it should assist them in their process.

3.3 PROJECT BRIEF

Our solution needs to be able to compute tumor probability maps from a WSI (Whole Slide Image). It needs to be robust against the different types of scanners used to scan the image as well as the different types of staining techniques that are used by pathologists. It also needs to compute these maps in a feasible amount of time as compared to a pathologist (30 hours). Furthermore, it should be able to be incorporated into a pathologist's pipeline for easy usage.

Chapter 4

METHODOLOGY

'CataNet' is computation tool to detect micro and macro-metastatic tissues from whole slide images (WSI) of breast sentinel lymph nodes. The project employs the help of several image processing techniques and state-of-the-art machine learning algorithms. More specifically, deep learning will be used to solve this problem. The tool consists of two major modules:

1. Slide level: Whole Slide Image classification module
2. Patient level: Pathologic lymph node classification (pN-stage) module

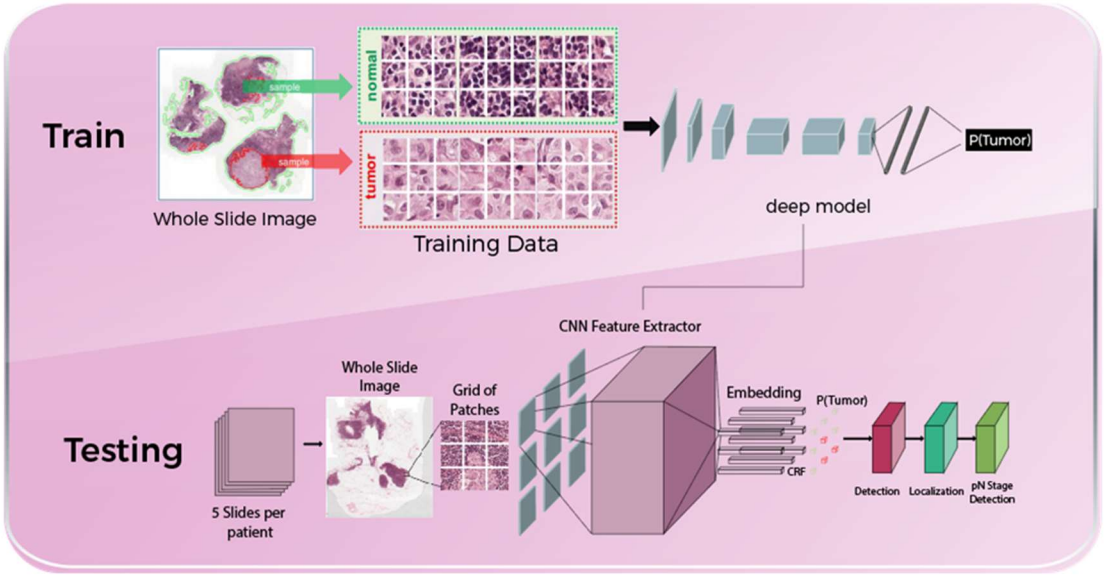

Figure 4: Pipeline for CataNet

4.1 SLIDE LEVEL: WHOLE SLIDE IMAGE CLASSIFICATION MODULE

In our project we tried to overcome the problems and challenges faced by classical methods with the development of Deep Learning based classification pipeline for finding metastatic breast cancer from digital whole slide images. This module is responsible for detection and localization of tumor in WSIs. 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

4.1.1 Region of Interest (ROI) detection with Image Processing

The first stage in our pipeline is to identify tissue region from whole slide image by excluding background white space. As explained before, WSIs are large GigaPixel (10^6 x 10^6) images, therefore it takes significant amount of time in order to process even a single image. Moreover, nearly 80% of WSI consists of the useless fat while only 20% forms the tissue region where the tumor may be present. So, ROI helps to reduce computation time significantly, as we only have to process the regions where probability of having tumor is high. Open-CV (Python API) is used to perform various operations involved in finding ROIs. This process has three steps:

4.1.2 Thresholding in RGB Space

Initially, Otsu's thresholding is used for thresholding in R, G and B space separately and then their logical AND is taken.

4.1.3 RGB to HSV Conversion

Then, the original image is converted from RGB color space to HSV (hue, saturation and value) color space. In HSV color space, color values are analyzed with greater convenience as values are more intuitive and easy to represent.

4.1.4 Binary Mask Generation

Once image is converted to HSV, next step is to create a binary mask by using Otsu thresholding in S component values. Now again a logical AND is taken between this mask and the mask previously created. The binary mask contains white pixels in areas where pixel values falls within the filtered range and black pixels everywhere else.

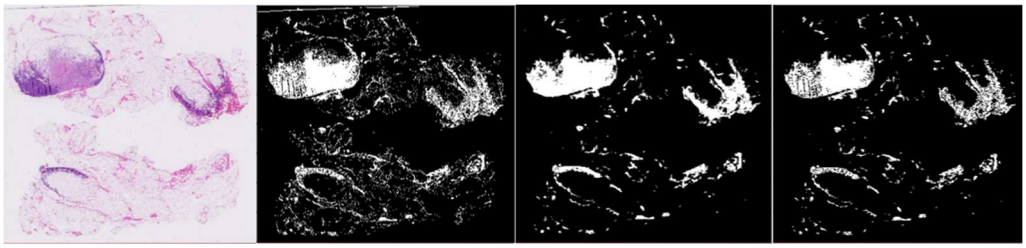

Figure 5: (From Left to Right) a) Original WSI b) Otsu Thresholding in RGB Spaces c) Otsu thresholding in Saturation space of HSV Model d) Logical AND of 'b' and 'c'

4.1.5 Constructing Training Data: Tiling ROI

After the detection of Region of Interest, training data needs to construct in order to train deep model. We have used breast cancer dataset which is available in form of whole slide images (WSIs) provided by Camelyon (the details of the dataset will be shared later in Section 6.1). Although the problem makes more sense with the image semantic segmentation task but due to large size of the dataset images we can't feed the whole image to our network. Also, it is not possible to load a full image into the computer memory, which makes it infeasible to analyze an entire whole image at once. Due to this

constraint, we decided to perform analysis on patches. We extracted thousands of patches from ROIs of each WSI randomly.

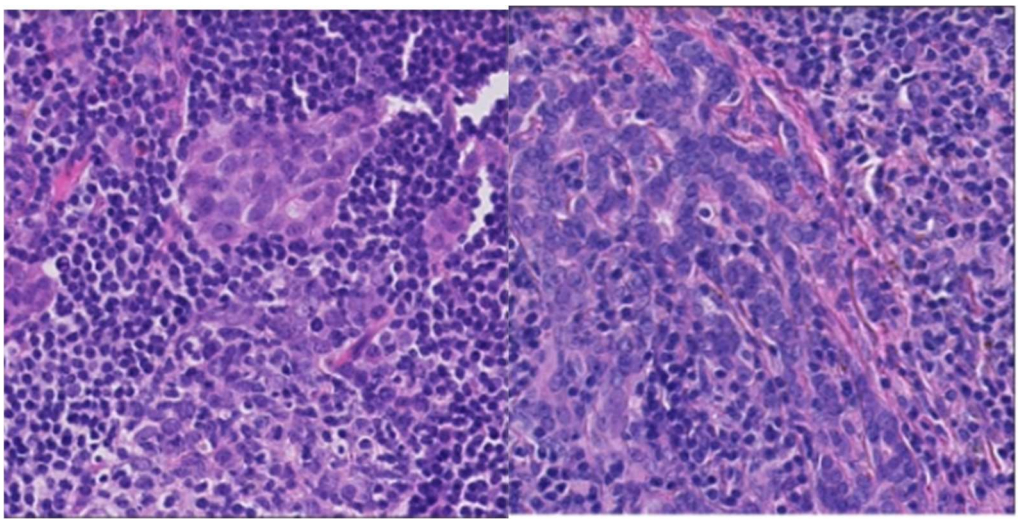

Figure 6: Extracted Normal and Tumor Patches

4.1.6 Training Deep ConvNet for Tile-Based Classification

Deep learning based models have shown state-of-the-art results in medical science challenges which prove the potential of applying deep learning techniques to solve real life health problems. On the other hand, conventional machine learning techniques require lots of manual steps for object detection, object segmentation and feature extraction. In this stage we train Deep ConvNet for patch based classification. We tried following architectures of CNN

1. Resnet'18
2. Resnet'18 with CRF
3. Inception-Resnet-v2 with CRF
4. CataNet with CRF

The trained model has the ability to distinguish between Tumor and Negative patches.

4.1.7 Building Tumor Probability Heat-Maps

After training Deep ConvNet model, next step is to generate tumor probability heat-maps for each WSI. This process was again performed in two steps:

4.1.8 Extraction of ROI

As explained earlier, ROI was extracted for each individual WSI, so that prediction should only be made for the tissue region and computation could be done efficiently.

4.1.9 Heat-Map Generation

Using trained deep Model, we obtained tumor probabilities for patches extracted from the ROI, and then embedded predictions of patches from individual WSIs into a single heat-map. In this way, 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 below demonstrates the step by step process for building a tumor probability heat-map.

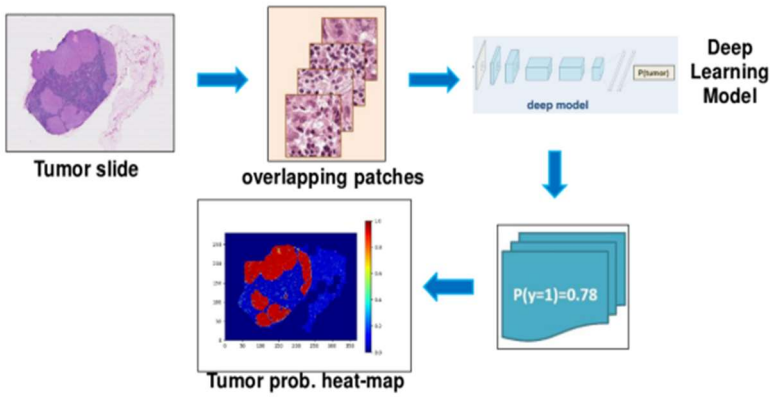

Figure 7: Pipeline for Heat-map Generation

4.1.10 Post-processing on heat-maps for slide-based classification

At slide-based classification, we perform the post-processing to perform the tasks of detection and localization of tumor.

4.1.11 Detection of tumor in WSI

In post-processing, we report the maximum tumor probability of a slide's heat-map as the slide's tumor probability. Detection of tumor is evaluated using ROC - AUC (receiver operating characteristic curve). Using the maximum value of each slide's heat-map, we achieved AUCs > 96%, statistically indistinguishable from the current best results.

4.1.12 Localization of tumor

We extract connected components from heat-maps and apply non maximum suppression on them. Maximum probability is reported along with the x-y coordinates. Localization of tumor is evaluated using FROC (free-response receiver operating characteristic). FROC will be explained with detail in section: 6.3.2.

	A	B	C
1	0.99987	32672	35104
2	0.99987	21088	24480
3	0.99983	36000	29984
4	0.99982	18976	21920
5	0.99977	72608	45728
6	0.99973	34784	35616
7	0.99972	27872	34272
8	0.99963	20256	28896
9	0.99962	22560	24672
10	0.99958	19040	28960
11	0.99956	25376	36000

Figure 8: Localization of Tumor Region

4.2 PATIENT LEVEL: PATHOLOGIC LYMPH NODE CLASSIFICATION (PN-STAGE) MODULE

This module is responsible for predicting the pN-stage for each patient. The pipeline followed by this module is given in figure.

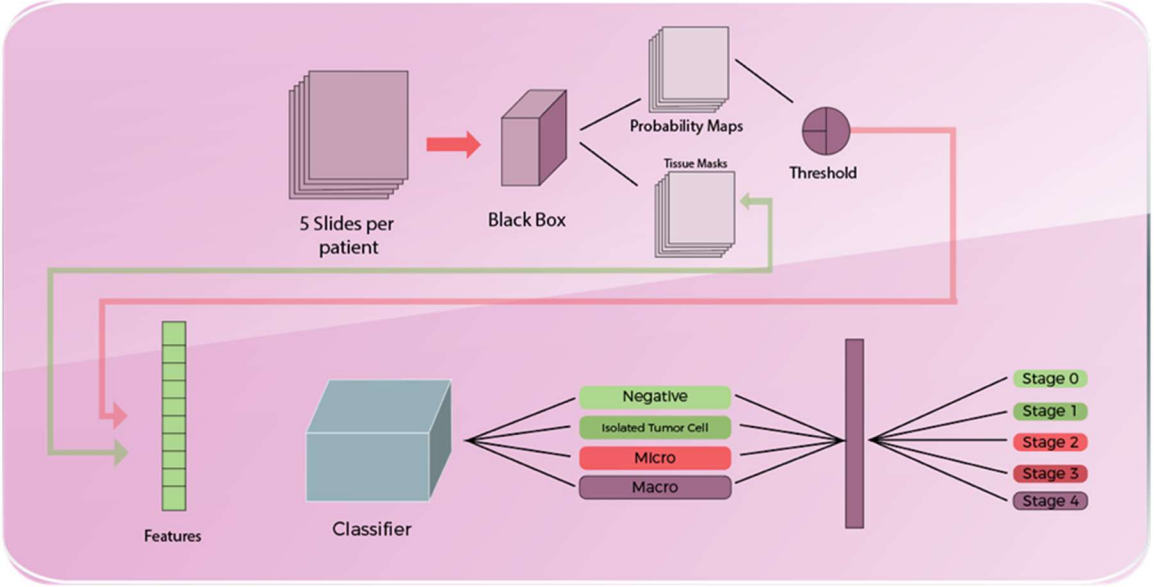

Figure 9: Pipeline for Patient Level Classification

Our patient level classification pipeline consists of five stages:

1. Extracting Tumor Maps and Tissue Masks
2. Extracting features
3. Train a classifier to predict metastasis type
4. Determining the pN stage

4.2.1 Extracting Tumor Maps and Tissue Masks

In the first step, tumor probability maps and tissue masks are extracted from the WSIs of each patient using the Slide-Level CNN module. Afterwards the maps are threshold at different levels from 0.3-0.9.

4.2.2 Extracting Features

After extraction of probability maps and tissue masks, different hand crafted features are extracted from the masks and probability maps at each threshold level.

4.2.3 Train a classifier to predict metastasis type

A classifier is trained to predict the slide-level metastasis level. The levels are specified as under:

a. Negative
b. Isolated Tumor Cells (ITCs)
c. Micro-metastasis
d. Macro-metastasis

We trained two classifiers to determine the tumor spread namely,

a. Random Forest Classifier
b. XGBoost Classifier

An ensemble of both of them was used. The details are mentioned in Section 6.

4.2.4 Determining the pN stage

Once we have slide-level metastasis level for each WSI of the patient, a pN-stage is assigned to the patient based upon a set of rules by pN-staging system as mentioned in table below.

Table 1: pN Stage Criteria

pN-Stage	**Criteria**
pN0	No micro-metastases or macro-metastases or ITCs found.
pN0(i+)	Only ITCs found.
pN1mi	Micro-metastases found, but no macro-metastases found.
pN1	Metastases found in 1–3 lymph nodes, of which at least one is a macro-metastasis.
pN2	Metastases found in 4–9 lymph nodes, of which at least one is a macro-metastasis.

Chapter 5

DETAILED DESIGN AND ARCHITECTURE

5.1 SYSTEM ARCHITECTURE

This section discusses detailed architecture and design of our system. The basic idea is to understand how the individual parts work together to provide the desired functionality. We have highlighted the main components, information flow and overall structure system.

5.1.1 Architecture Design Approach

We are using Pipeline approach in our project. Pipelining is a method where there are stages and these stages are connected with one another to form a pipe like structure. The flow of data in this pipeline is mentioned in figure given below.

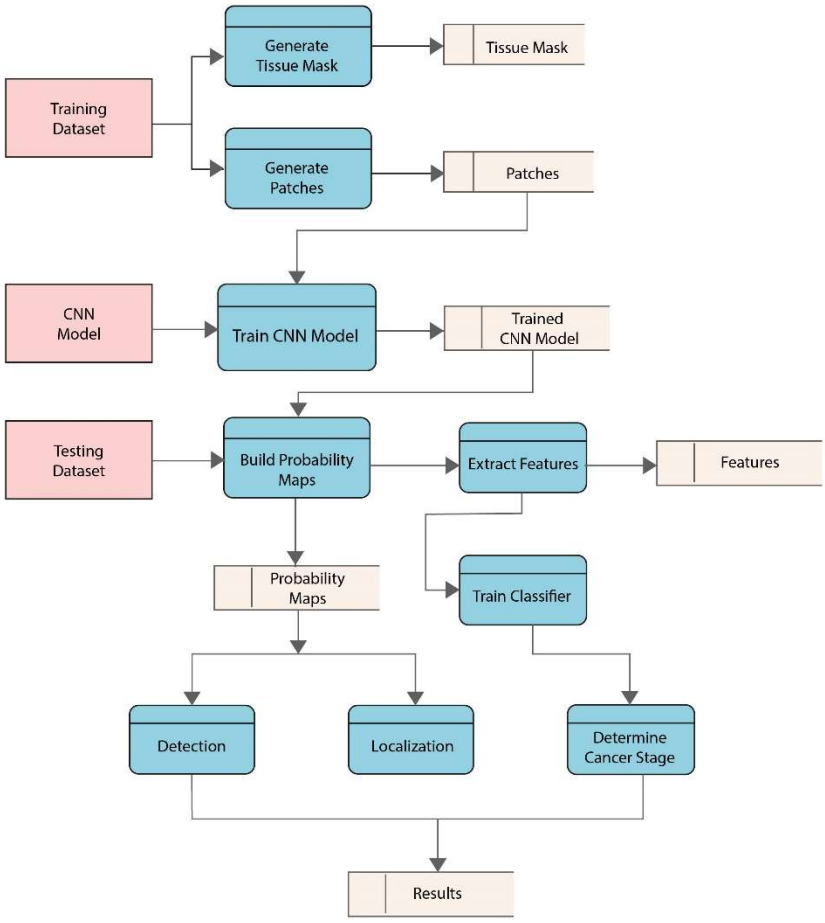

Figure 10: CataNet Data Flow Diagram

5.1.2 Architecture Design

The overall system is divided into these two major modules.

5.1.3 Slide-Level: Detection and Localization of Cancer Cells

The pipeline to be followed is as under:

a. Region of Interest (ROI) detection with Image processing

b. Construct training data: Extract Positive & Negative tiles from ROI

c. Train Deep Convolutional model for tile-based classification

d. Building tumor probability heat-maps using trained model

e. Post-processing on heat-maps for slide-based classification

At the end of phase 1, given a hematoxylin and eosin (H&E) stained whole-slide images of lymph node, the product classifies whether the image contains tumor or not, and if present then localizes it within the image.

5.1.4 Patient-Level: Determining the pathologic N-stage (pN-stage) label per patient

The pipeline in this stage comprises a separate model for detecting the stage of cancer in patient.

1. Extracting Tumor Maps and Tissue Masks
2. Extracting features
3. Train a classifier to predict metastasis type
4. Determining the pN stage

At the end of phase 2, given at least 5 whole-slide images of a patient, the product determines the pN-stage of the patient as provided by the Union for International Cancer Control (UICC). The overall structure of the system and dependencies in the system is further explained in the Package Diagram in Figure 11.

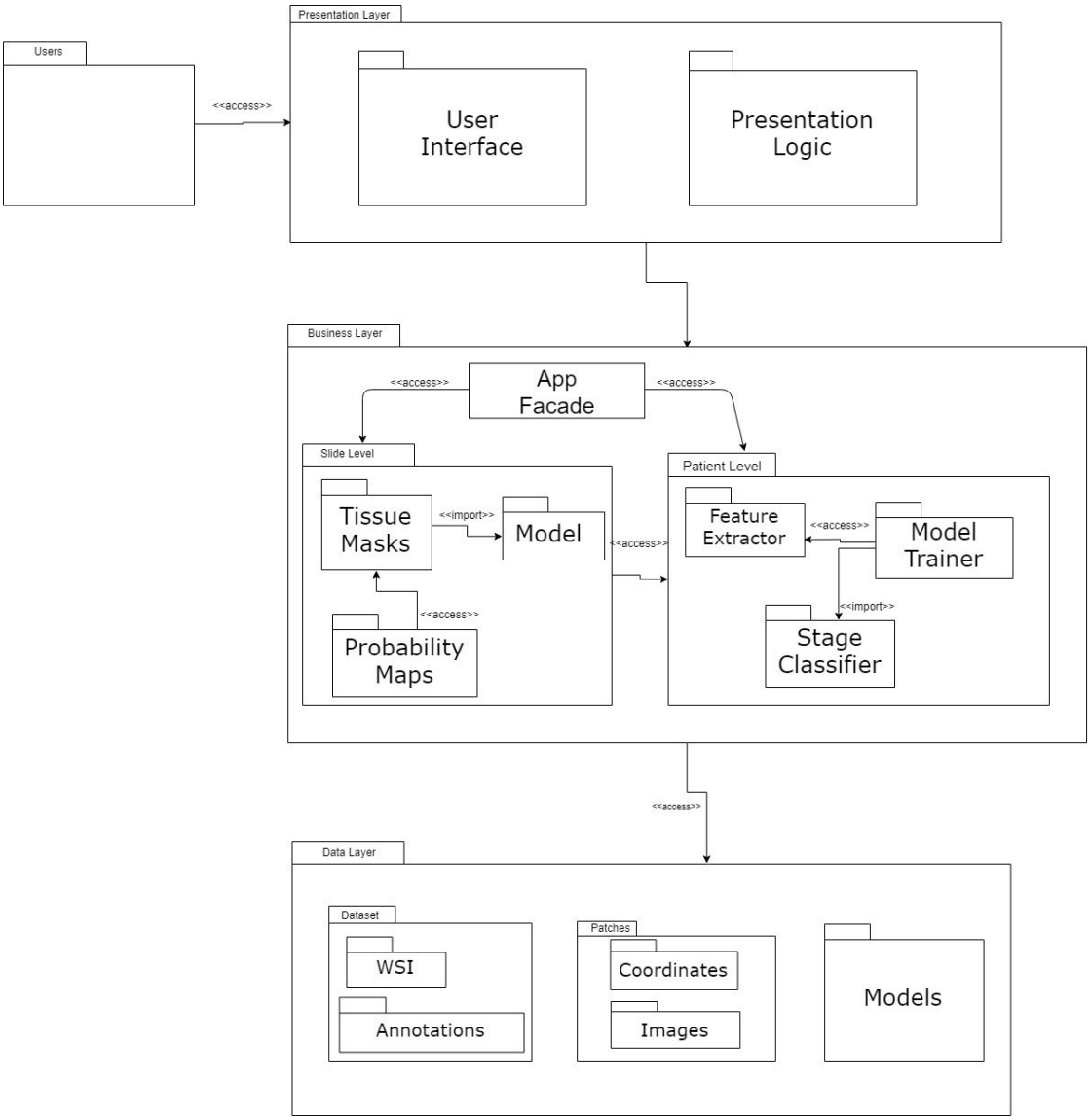

Figure 11: CataNet Package Diagram

5.1.5 Subsystem Architecture

The system is divided into two major components and decomposition of these components can be observed through the components diagram in Figure 12.

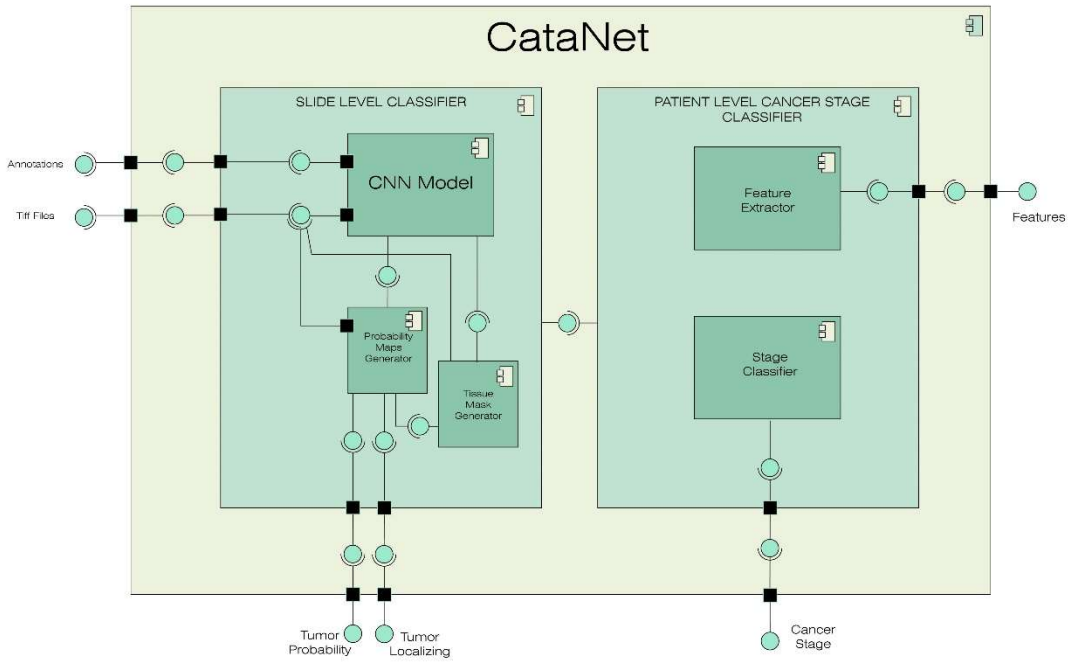

Figure 12: CataNet Component Diagram

5.2 DETAILED SYSTEM DESIGN

Since the system is comprised of two major modules, the detailed design will be discussed for these two modules as well. Each module will cover all the functionalities listed in it.

5.2.1 Classification

The system consists of two major components as mentioned below.

- Slide Level Classifier

  The slide level classifier is again divided into two sub components:

  - Patch Classifier
  - WSI Classifier
- Patient Level Cancer Stage Classifier

We will explain each component in detail below.

5.2.2 Slide level classifier

5.2.2.1.1 Definition

This component classifies a slide as tumor or normal by interacting with a trained convolutional neural network.

5.2.2.1.2 Responsibilities

The primary responsibilities of this component are given as follows:

- To allow user input the slide that needs to be classified as tumor or normal.
- Input the slide in a trained neural network.
- Show the results.

5.2.2.1.3 Constraints

To perform above mentioned task, there must be a trained deep neural network that can classify the slide. Additionally, an internet connection is also required in order to upload the slide.

5.2.2.1.4 Uses/Interactions

This component uses convolutional neural network directly. User doesn't interact with ConvNet to classify a slide. User only needs to upload a slide.

5.2.2.1.5 Resources

The resources that this component requires are:

- A computer system with a GPU GTX 1080 Ti or higher. The system will deploy and use CUDA library to communicate with the GPU.
- 16GB or more built-in storage.
- 12 Cores CPU or higher.
- 1TB Storage memory or higher
- Trained Convolutional Neural Network

5.2.2.1.6 **Interface/Exports**

As this part of the system is for the end user of the system, the Client side of the System is therefore created as a User friendly interface with details hidden from the end user.

5.2.3 Patch based classifier

5.2.3.1 Definition

This component classifies a single patch of a given slide image as a tumor or a normal patch by interacting with a trained convolutional neural network.

5.2.3.2 Responsibilities

The primary responsibilities of this component are given as follows:

- To allow user to upload the slide whose patches needs to be classified as tumor or normal.
- Extract the patch from a slide.
- Input the patch in a trained neural network.
- Show the classification results for the patch.

5.2.3.3 Constraints

To perform above mentioned task, there must be a trained deep neural network that can classify the patch. Additionally, an internet connection is required in order to upload the slide.

5.2.3.4 Uses/Interactions

This component uses convolutional neural network directly. User doesn't interact with ConvNet to classify a patch. User only needs to upload a slide.

5.2.3.5 Resources

The resources that this component requires are:

- A computer system with a GPU GTX 1080 Ti or higher. The system will deploy and use CUDA library to communicate with the GPU.
- 16GB or more built-in storage.
- 12 Cores CPU or higher.
- 1TB Storage memory or higher
- Trained Convolutional Neural Network

5.2.3.6 Interface/Exports

As this part of the system is for the end user of the system, the Client side of the System is therefore created as a User friendly interface with details hidden from the end user.

5.2.4 WSI Classifier

5.2.4.1 Definition

This component classifies a whole slide image as a tumor or a normal by interacting with a trained convolutional neural network.

5.2.4.2 Responsibilities

The primary responsibilities of this component are given as follows:

- To allow user to upload the WSI that needs to be classified as tumor or normal.
- Extract the patches from a WSI image.
- Use the patch based classifier to generate a heat map for the WSI
- Show the classification results in terms of detection and localization.

5.2.4.3 Constraints

To perform above mentioned task, there must be a trained deep neural network that can classify the patch. An internet connection is required to upload slide. Moreover, size of WSI is very large therefore it is time consuming task.

5.2.4.4 Uses/Interactions

This component uses convolutional neural network directly. User doesn't interact with ConvNet to classify a WSI. User only needs to upload a whole slide image and this component will show the detection and localization results in terms of tumor probability and heat map.

5.2.4.5 Resources

The resources that this component requires are:

- A computer system with a GPU GTX 1080 Ti or higher. The system will deploy and use CUDA library to communicate with the GPU.
- 16GB or more built-in storage.
- 12 Cores CPU or higher.
- 1TB Storage memory or higher
- Trained Convolutional Neural Network

5.2.4.6 Processing

A whole slide image is given as input to a trained deep model through the user interface. In backend the window is convolved over WSI to classify the patches as tumor or normal to generate probability heat-map for WSI. Based on the generated heat-map, results are shown to the pathologist.

5.2.4.7 Interface/Exports

As this part of the system is for the end user of the system, the Client side of the System is therefore created as a User friendly interface with details hidden from the end user.

5.2.5 PATIENT LEVEL CANCER STAGE CLASSIFIER

5.2.5.1 Definition

Given 5 whole slide images of a patient, this component determines the stage of cancer, from which patient is suffering. The details of the stages are mentioned in Table 1.

5.2.5.2 Responsibilities

The primary responsibilities of this component are given as follows:

- Allow user to upload 5 WSIs of a patient
- Feed each slide in a trained deep model and generate heat-maps.
- Report maximum value of each slide's heat-map
- Determine the stage of a cancer.

5.2.5.3 Constraints

To perform above mentioned task, there must be a trained deep neural network that can classify the patch. An internet connection to upload WSIs. Moreover, generating heat-maps for each slide is time consuming task.

5.2.5.4 Uses/Interactions

This component uses convolutional neural network directly. User doesn't interact with ConvNet to classify a patch. User only needs to upload a slide.

5.2.5.5 Resources

The resources that this component requires are:

- A computer system with a GPU GTX 1080 Ti or higher. The system will deploy and use CUDA library to communicate with the GPU.
- 16GB or more built-in storage.
- 12 Cores CPU or higher.
- 1TB Storage memory or higher
- Trained Convolutional Neural Network

5.2.5.6 Processing

Five whole slide images per patient are fed to a trained deep model through the user interface. In backend the window is convolved over WSIs to classify the patches as tumor or normal to generate probability heat-map for WSI. Based on the features of generated heat-maps, cancer stages are determined and results are shown to the pathologist.

Chapter 6

IMPLEMENTATION AND TESTING

6.1 DATASET

In our project, we used a breast cancer dataset as part of Camelyon challenge. The dataset is divided in two parts

- Camelyon'16
- Camelyon'17

6.1.1 Camelyon'16

Camelyon'16 contains a total of 400 whole slide images (WSIs) of sentinel lymph node from two independent datasets collected in Radbound University Medical Center (Nijmegan, the Netherlands), and the University Medical Center Utrecht (Utrecht, the Netherlands). the dataset contains:

- Training Slides
    - Normal Slide: 159
    - Tumor Slide: 111
- Testing Slides
    - Annotated Slides: 130

The ground truth data for containing metastasis is provided in two formats:

- XML files containing vertices of the annotated contours
- Binary Mask

Table 2: Training Slides in Camelyon'16

Source	Train Tumor	Train Normal
Radbound UMC	70	100
UMC Utrecht	41	59
Total	111	159

Table 3: Testing Slides in Camelyon'16

Source	Annotated Slides
Radbound UMC	80
UMC Utrecht	50
Total	130

6.1.2 Camelyon'17

The data set for CAMELYON17 is collected from 5 medical centers in the Netherlands. Whole slide images are provided as TIFF images. Lesion-level annotations are provided as XML files. For training, data of 100 patients is given and for testing we data of another 100 patients is also provided. Since we have 5 slides per patient, therefore total number of slides is 1000.

Table 4: Slides in Camelyon'17

	Annotated Slides
Training Slides	500
Testing Slides	500
Total	1,000

6.2 EXPERIMENT SETUP

To successfully perform experiments on proposed system following requirements were setup.

6.2.1 Hardware Configurations

- A computer system with a GPU GTX 1080 Ti or higher. The system will deploy and use CUDA library to communicate with the GPU.
- 16GB or more built-in storage.
- 12 Cores CPU or higher.
- XBOX kinect
- 1TB Storage memory or higher

6.2.2 Software Interfaces

- Python 3.6
- PyTorch (0.3.1)/CUDA 9.0
- SciPy (1.0.1)
- YOLOv3
- Matplotlib (2.2.2)
- TensorboardX
- PIL (5.1.0)

6.3 EVALUATION METRICS

6.3.1 ROC Curve (AUC)

Receiver operating characteristic (ROC) analysis is performed at the slide level and the measure used for comparing the algorithms is the area under the ROC curve (AUC). ROC is a probability curve and AUC represents degree or measure of separability. It tells how much model is capable of distinguishing between tumor or normal slides. The ROC curve is plotted with True Positive Rate (TPR) against the False Positive Rate (FPR) where TPR is on y-axis and FPR is on the x-axis.

6.3.2 FROC

Free Response Operating Characteristic (FROC) curve is used to evaluate the performance of the algorithms for lesion detection/localization. It is similar to ROC analysis, except that the false positive rate on the x-axis is replaced by the average number of false positives per image. In FROC, we consider a true positive, if the location of the detected region is within the annotated ground truth lesion.

- If there are multiple findings for a single ground truth region, they will be counted as a single true positive finding and none of them will be counted as false positive.
- All detections that are not within a specific distance from the ground truth annotations will be counted as false positives.

6.3.3 KAPPA

Cohen's kappa coefficient (κ) is a statistic which measures inter-rater agreement for qualitative (categorical) items. It is generally thought to be a more robust measure than simple percent agreement calculation, as κ takes into account the possibility of the agreement occurring by chance.

6.4 DEEP LEARNING MODELS

For our project we have utilized four models.

- ResNet-18-CRF
- Inception ResNet CRF
- ResNet-18
- CataNet-CRF

6.4.1 Inception ResNet CRF

We used Inception ResNet with additional layer of CRF to distinguish between tumor and normal patch. The CRF layer was attached at the end of the layers of the actual CNN model and it was fed with the embeddings as well as the label of the patches (belonging to a grid of 3x3). There were total 468 layers including CRF layer with 1536 embeddings. The CRF layer is explained further in another model.

Schema for original Inception-ResNet-v2 network is given in figure below as mentioned in Google paper.

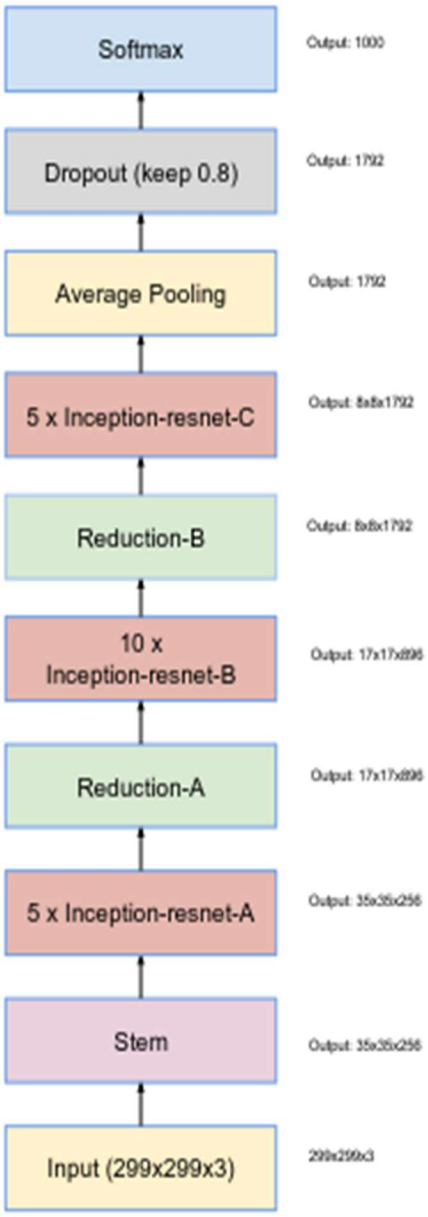

Figure 13: Schema for Inception ResNet v2

6.4.2 ResNet 18

We also used Resnet-18 due to its less parameters as compared to its other variations. The major problem solved by ResNet is that when deeper networks start converging, a degradation problem occurs. With the network depth increasing, accuracy gets saturated and then

degrades rapidly. Here comes the ResNet model which introduces the skip connections.

An abstract schema is given in figure below.

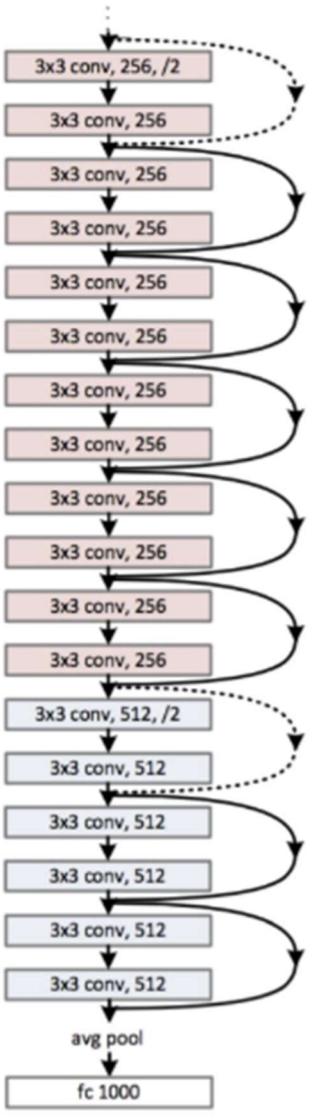

Figure 14: Schema for ResNet'18

6.4.3 ResNet 18-CRF

We also tested ResNet with additional layer of CRF to distinguish between tumor and normal patch. The CRF layer was attached at the end of the layers of the actual CNN model and it was fed with the embedding as well as the label of the patches (belonging to a grid of 3x3). There were total 19 layers including CRF layer with 512 embedding.

6.4.4 CataNet-CRF

We tested our own architecture inspired by a model used to detect the left vertebrate of the heart. A CRF layer was similarly applied to this model. The model without the CRF layer is shown below:

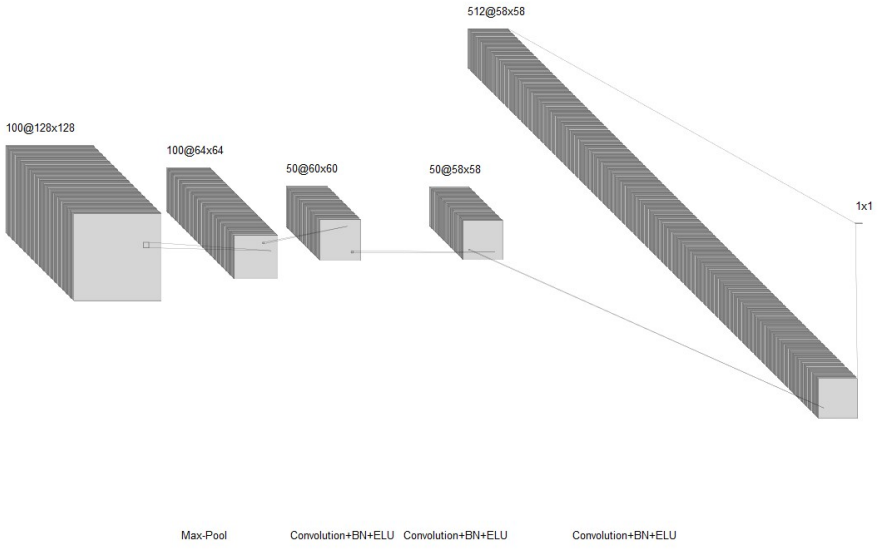

Figure 15: Design of CataNet

6.4.5 Augmentations

The pixel values of the patches were normalized by subtracting 128 and dividing 128. Color jitter was also added with the following configuration:

- Brightness: 64/255
- Hue: 0.04
- Saturation: 0.25
- Contrast: 0.75

Random left/right flip was also performed with a probability of 0.5. Random rotation was also applied in multiples of 90 degrees.

6.4.6 CRF Layer

The CRF layer uses a grid size of 3x3 (total of 9 patches within the image). The CRF layer uses the feature embedding (from fully connected layers) and their corresponding labels from the final layer of a CNN. It uses these feature functions and labels to map the relationship in a 3x3 grid within the original patch. CRFs predict structure and sequence in a given input. In basic terms, a CRF uses the sum of unary potentials and pairwise potentials between indexes in order to calculate the probability of a particular patch being assigned a particular label. The unary potential shows how likely an input is to be labeled as tumor on non-tumor whereas the pairwise potential tells you how likely it is to be labeled as tumor or non-tumor depending on whether the previous instance was tumor or non-tumor respectively. In order to ensure that this is a valid probability distribution, the resultant probabilities are divided by the sum of all probabilities over all possible labels known as a partition function. This normalizes each probability.

In our particular case, in order to find out how likely it is that the current index is assigned the same label as the previous one; we use the cosine similarity index between the feature functions. The CRF is rewarded for assigning them the same labels if the distance is less otherwise it is penalized for assigning them different labels.

Table 5: Model Details

Architecture	Number of Layers	Embeddings	Parameters
Inception ResNet v2-CRF	467+1	1536	55.8 Million
ResNet-18	18	512	11.7 Million
ResNet-18-CRF	18+1	512	11.7 Million
CataNet-CRF	13+1	512	1.9 Million

6.5 TRAINING AND TESTING OF SLIDE LEVEL CLASSIFIER

6.5.1 ROI Detection

As the size of our slides was very large therefore it was not feasible to directly feed it into our CNN. Instead, we had to extract tumor/non-tumor patches from each slide and then train our CNN on them. In order to extract the patches by random sampling, we had to remove the non-tissue region from the image. This was done by using Otsu's thresholding method in the saturation space from HSV as well as in RGB and then using logical AND on both of the masks. The resulting mask is shown in figure 15:

Figure 16: Tissue Mask from WSI

6.5.2 Constructing Training Data

Training data is constructed by extracting 1023 x 1023 patches from the tissue region of the WSIs. A patch is labeled as tumor if it contains even one pixel of tumor. We used a pre-sampled set of coordinates to construct our first dataset and separately constructed a randomly sampled second dataset. Augmentations were also used to increase the quality of the data available. The details of augmentation were mentioned in section 6.4.5. The details of both sets are shown below:

Patch Dataset 1	
Label	No. of Patches
Normal	200,000
	20,000
Tumor	200,000
	20,000

Figure 17: Dataset 1

Patch Dataset 2	
Label	No. of Patches
Normal	900,000
	90,000
Tumor	300,000
	30,000

Figure 18: Dataset 2

6.5.3 Training Deep ConvNet

Our training setup was on Google Cloud and was as follows:

- CPU: 16 x 2.30GHz Intel Xeon vCPUs
- GPU: 1 x NVidia P100
- RAM: 16GB Memory
- HDD: 10 TB

The hyper parameters for each model are given in table 6:

Table 6: Hyperparameters for the Models

Model	**Resnet-18 CRF**	**Inception ResNet v2 CRF**	**CataNet-CRF**
Batch Size	32x2	2x2	6x4
Image Size	768	1023	1023
Patch Size	256	341	341
Crop Size	224	299	256
Learning Rate	0.001	0.001	0.001
Momentum	0.9	0.9	0.9

6.5.4 Testing

Patches were extracted from the WSI with a stride of 64 at level 0. Each patch was classified with a probability of being tumor or not. Once all the probability maps were generated, we calculated the AUC and FROC for all the probability maps.

6.6 PATIENT LEVEL CANCER STAGE CLASSIFIER

6.6.1 Feature Extraction

In order to train patient level classifier, we generated tissue masks and probability maps for each WSI of a patient. The tissue masks were generated at level 6 of WSIs and probability maps were generated using Resnet-18 with CRF as CNN module in Patch Level Classifier. Then we threshold the probability maps at levels of 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 and 0.9. After that, at each threshold level we

extracted features as mentioned in the table 7. These features were hand crafted and extracted through testing their efficiency on a validation set.

Table 7: Extracted Features

S. No.	Feature
1	Largest region's major axis length
2	Largest region's maximum confidence probability
3	Largest region's average confidence probability
4	Largest region's area
5	Average of all region's averaged confidence probability
6	Sum of all region's area
7	Maximum confidence probability in WSI
8	Average of all confidence probability in WSI
9	Number of regions in WSI
10	Sum of all foreground area in WSI
11	Foreground and background area ratio in WSI

6.6.2 Predicting pN Stage

Once features have been extracted we used an ensemble of Random Forest Classifier and XGBoost Classifier. Table 8 shows a fair comparison of the two classifiers. The ensemble helps to tackle the case of class imbalance problem of dataset mentioned in table. The Random Forest Classifier was efficient in correctly labeling Negative Slide while XGBoost was efficient with the ITCs, micro and macro

metastasis. So the ensemble was used in such a way that in case of conflict between the two classifiers if Random Forest is reporting "Negative" while XGBoost is reporting "ITC, Macro or Micro", the ensemble would report "ITC, Macro or Micro" as XGBoost knows better about them while Random forest is prone to reporting "Negative" class in most cases. Also if Random Forest is reporting "ITC, Macro or Micro" while XGBoost is reporting "Negative", the ensemble would report "ITC, Macro or Micro".

Table 8: Difference between Random Forest and XGBoost

Random Forest	**XGBoost**
Every sample had the same probability to appear in a new dataset.	Samples were weighted so some of them appeared more often
Reduced the variance	Reduced the bias
Efficient in correctly labeling Negative slides	Efficient in correctly labeling "ITCs, Macro and Micro metastasis" slides

Table 9: Class Distribution

Label	No. of Sample Slides
Negative	318
ITCs	26
Micro-Metastasis	59
Macro-Metastasis	87

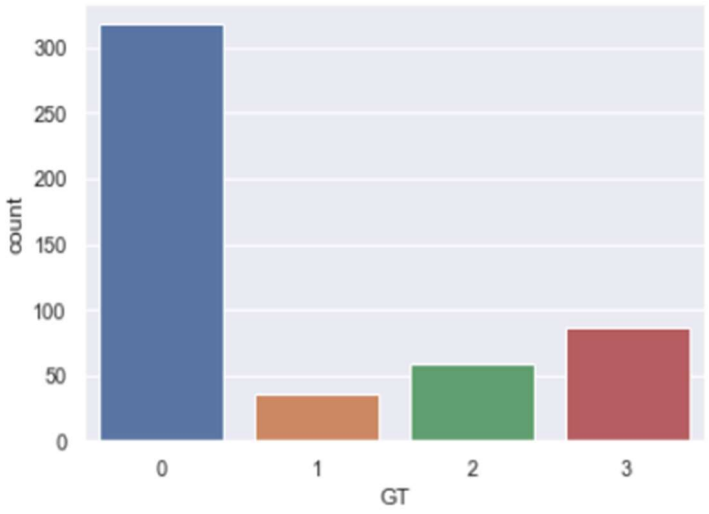

Figure 19: Visualization of Class Distribution

Chapter 7

RESULTS AND DISCUSSION

In this section we briefly present the evaluation results of the tests on Datasets of Camelyon'16 and Camelyon'17.

7.1 SLIDE LEVEL CLASSIFIER

Our results for each model are shown below. Although, Inception-ResNet-v2-CRF gives us the best results, it took much longer to train than ResNet therefore we dropped it for ResNet-18-CRF which gave comparative validation accuracy.

Table 10: Resulting Metrics for each Model

Slide Level Classifier			
Architecture	**Patch Training Accuracy**	**ROC (Detection)**	**FROC (Localization)**
ResNet-18	92.42%	0.9644	0.7966
ResNet-18-CRF	92.96%	0.9668	0.8075
ResNet-18-CRF+RF	92.96%	0.9792	0.8075
Inception ResNet v2 CRF	95.8%	-	-

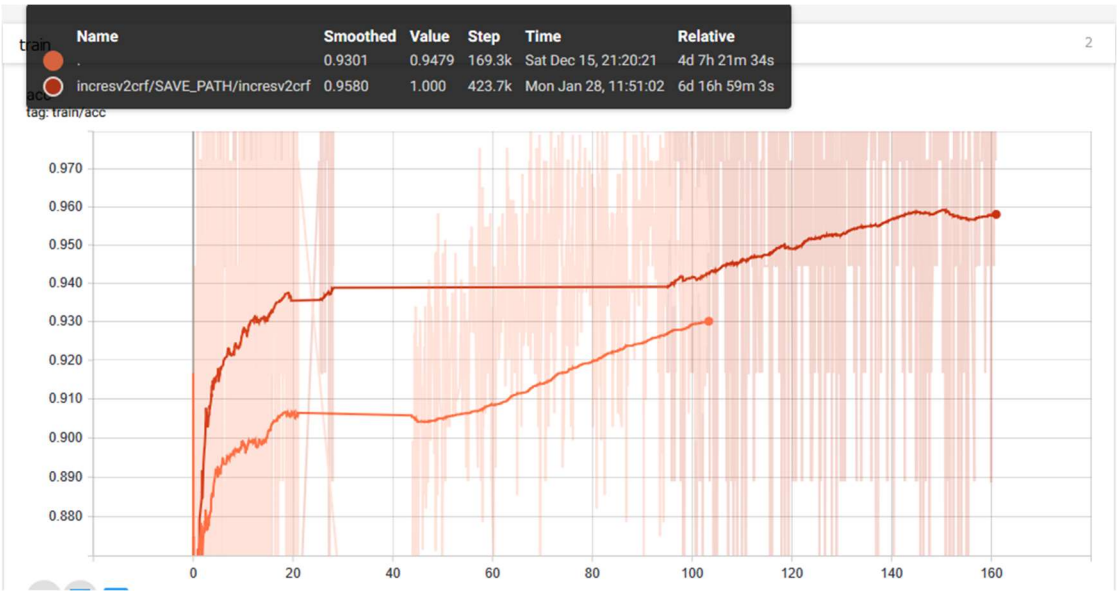

Figure 20: Comparison of IncResV2 (labeled as ".") and IncResV2CRF

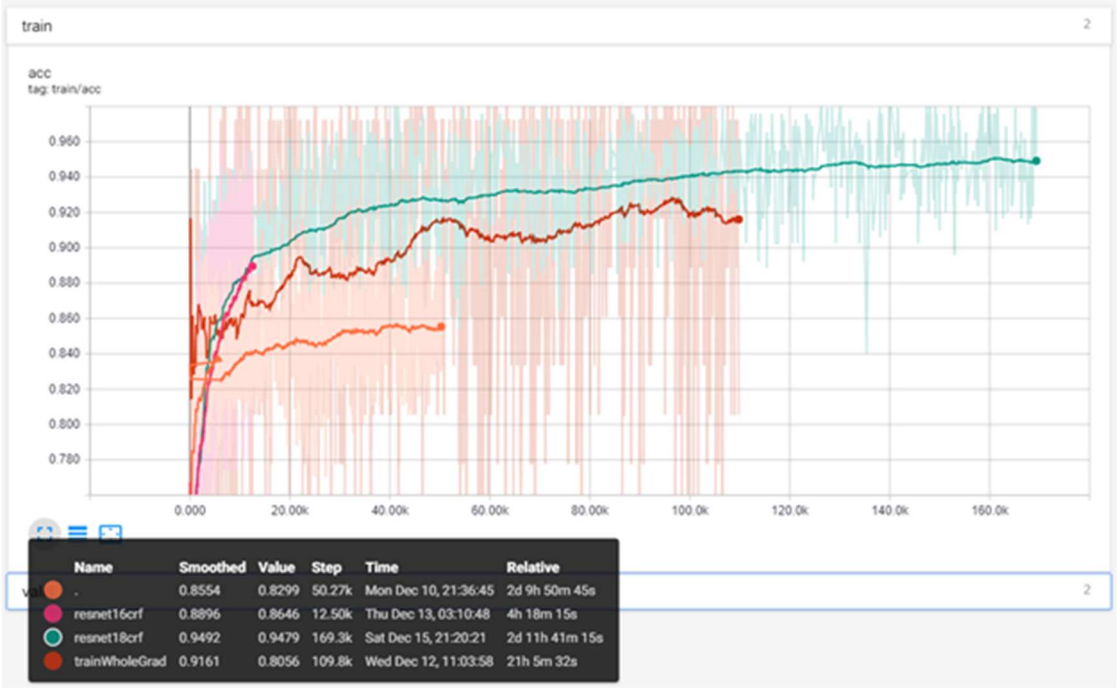

Figure 21: Comparison of ResNet18-CRF with some other models

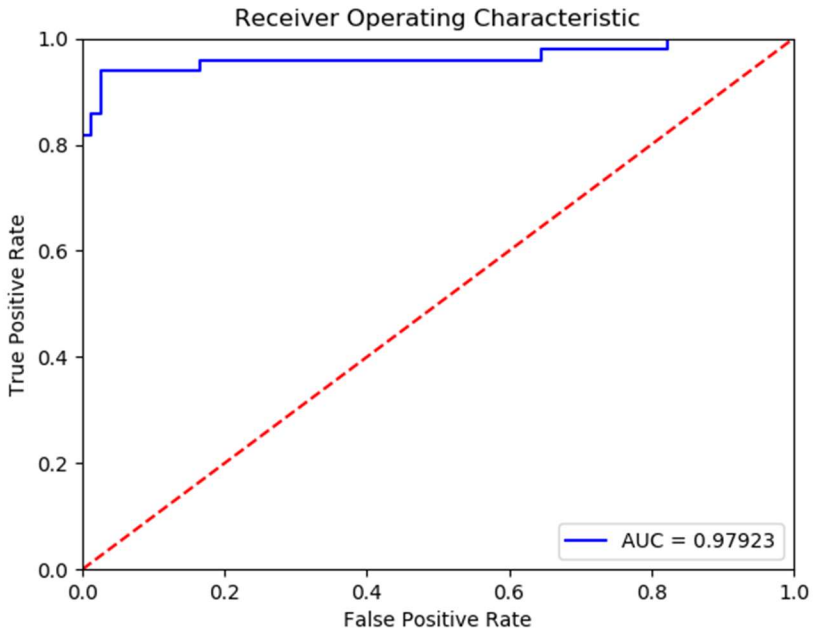

Figure 22: Visualization of ROC

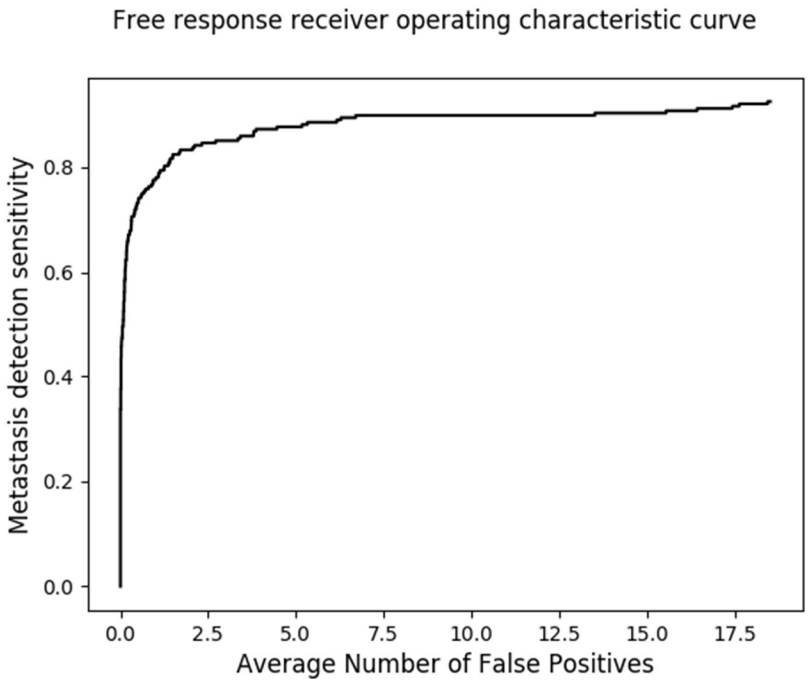

Figure 23: Visualization of FROC

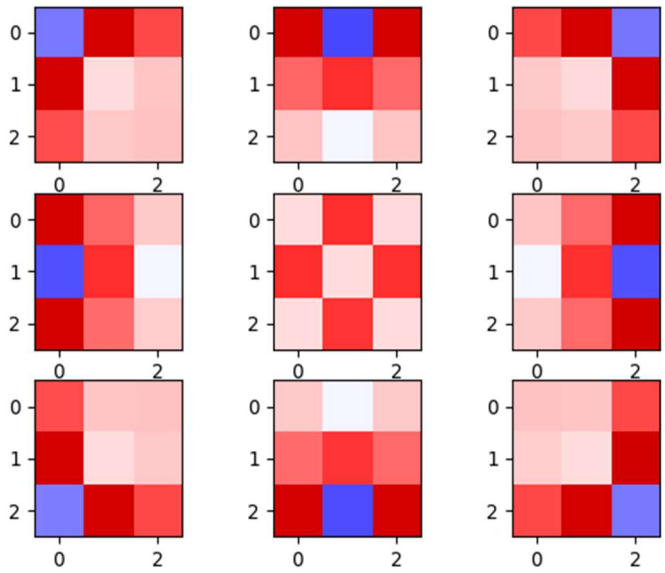

Figure 24: Visualization of CRF's Weights

7.2 PATIENT LEVEL CANCER STAGE CLASSIFIER

With using only Random Forest Classifier, we got Kappa score of 0.8345 while using only XGBoost gives us accuracy of 0.8075. This was improved to 0.8749 by using the ensemble as mentioned in Section 6.

Table 11: Patient Level Cancer Stage Classifier (using ResNet-18 as CNN Module)

Classifier	**Kappa (Localization)**
Random Forest	0.8345
XG Boost	0.8075
Ensemble	0.8749

Chapter 8

CONCLUSION AND FUTURE WORK

The aim of this report is to present a deep learning based system for detection of metastatic cancer from whole slide images of sentinel lymph nodes. The main challenges we faced were to improve training set to enhance previous system by avoiding misclassification of the normal lymph node region as cancer. Other important features of the system were to develop state-of-the art deep learning architecture to classify small patches from the large whole slide image and by carefully designing the post-processing methods to perform classification on slide level.

Classical methods were mainly focused on image analysis tasks such as color normalization, nuclear segmentation and feature extraction. Generally, image analysis alone is not enough for classification of the tumor, so in practice after image analysis; machine based classification models such as SVM (support vector machine) and Random Forest are required for end-to-end feature classification task. Deep learning approaches have performed exceptionally well 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. Unlike 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). Our project utilizes Inception Resnet CRF and Resnet 18 CRF and we obtained near human-level classification performance on the breast cancer test data. Although the performance of pathologist alone is currently superior to deep learning system alone, combining deep learning with 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 in medical domain can increase the speed, accuracy, reliability of diagnoses.

For improving the Slide Level Cancer Stage accuracy, we recommend following approaches

1. Sampling a higher number of patches from each WSI.

2. Sampling more non-tumor patches than tumor patches in order to reduce false positives.

3. Repetitive hard mining.

4. Up-scaling last two layers to achieve a label for each pixel (gives more detailed probability maps).

For improving the Patient Level Cancer Stage accuracy, we recommend following approaches

1. The first step is to increase the class imbalance problem by increasing the samples of "ITC, Micro-metastasis and Macro-Metastasis" class. In this way the classifier would become more efficient.

2. Another approach is to increase the number of centers from which the dataset is sampled. We found out that including the dataset from different center increases the Kappa Score as large as 20%. It also increases the robustness of our classification model.

3. We also observed that our model was having difficulties in identifying ITCs due to their smaller size as mentioned in Table. This problem can be solved by generating probability maps at a lower level than 6 although this would sufficiently increase the computation time.

Our future research direction could be to train and evaluate the performance of our system on some other cancer data set. It can be a large scale open source data such as TCGAs lung cancer dataset. Another important step would be to integrate staining normalization process in our classification pipeline in the future, because, it is very crucial part of building a generalize deep learning model for identifying metastatic regions as it removes the stain variability in WSIs which is induced due to different techniques used for staining. Stain normalization process has a great potential to enhance the ability of a deep model to classify, which shows that with stain normalization component on board, CataNet could produce even better results and could be extended to any kind of breast cancer dataset.

Chapter 9

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.
2. Y. Liu, K. Gadepalli, M. Norouzi, G. E. Dahl, T. Kohlberger, A. Boyko, S. Venugopalan, A. Timofeev, P. Q. Nelson, G. S. Corrado, et al. Detecting cancer metastases on gigapixel pathology images. arXiv preprint arXiv:1703.02442, 2017.
3. L. Hou, D. Samaras, T. M. Kurc, Y. Gao, J. E. Davis, and J. H. Saltz. Patch-based convolutional neural network for whole slide tissue image classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2424–2433, 2016.
4. Y. S. Vang, Z. Chen, and X. Xie. Deep learning framework for multi-class breast cancer histology image classification. arXiv preprint arXiv:1802.00931, 2018.
5. W. Zhu, Q. Lou, Y. S. Vang, and X. Xie. Deep multi-instance networks with sparse label assignment for whole mammogram classification. In International Conference on Medical Image Computing and Computer-Assisted Intervention, pages 603–611. Springer, 2017.
6. W. Zhu et al. Adversarial deep structured nets for mass segmentation from mammograms. In IEEE ISBI, 2018.
7. Li, Y., & Ping, W. (2018). Cancer Metastasis Detection With Neural Conditional Random Field. *CoRR, abs/1806.07064*.