REPUBLIC OF TURKEY
YILDIZ TECHNICAL UNIVERSITY
DEPARTMENT OF COMPUTER ENGINEERING

SEGMENTATION OF TUMOR REGION IN HISTOPATHOLOGY IMAGES

18011061 — Emre ARSLANOĞLU
19011913 — Muhanad TUAMEH

SENIOR PROJECT

Advisor
Prof. Songul VARLI

June, 2023

ACKNOWLEDGEMENTS

We express our sincere gratitude to all those who contributed and assisted in completing this project. Especially, we would like to thank our mentor Prof. Songül VARLI for her time, advices and valuable guidance.

We would like to extend our sincere gratitude to all the organizers of the PAIP challenge for preparing the data and providing it to us. It would be very hard to complete this project without the provision of this data.

Also, thanks to all my family members and friends for their love and support throughout our life. Their presence and encouragement have been a constant source of strength and motivation, inspiring us to overcome challenges and pursue our goals.

Emre ARSLANOĞLU
Muhanad TUAMEH

TABLE OF CONTENTS

LIST OF SYMBOLS

Ai	Activities of Daily Life
c	Alternate Step Test
C	Body Mass Index
CR	Cross Step moving on Four Stops
$fc(.)$	Dynamic Bayesian Networks
ΔH	Demura's Fall Risk Assessment Chart
λi	Electromyography
Ω	Faculdade de Engenharia da Universidade do Porto

LIST OF ABBREVIATIONS

PAİP	The Pathology AI Platform
HCC	Hepatocellular Carcinoma
SNUH	Seoul National University Hospital
SNUBH	Seoul National University Bundang Hospital
BMC	Boramae Medical Center
LiTS	The Liver Tumor Segmentation
TUM	Technical University of Munich
CT	Computerized Tomograph
FCN	Fully Convolutional Network
IoU	Intersection over Union
WSI	Whole Slide Images
GPU	Graphics Processing Unit
CPU	Central Process Unit
TL	Turkish Lira

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

SEGMENTAT1ON OF TUMOR REGION IN HISTOPATHOLOGY IMAGES

Emre ARSLANOĞLU
Muhanad TUAMEH

Department of Computer Engineering
Senior Project

Advisor: Prof. Songul VARLI

This study examines the segmentation of tumor patches within whole-slide images (WSI) and presents the findings from 3 different modeling approaches; CNN and ResNet as binary classifiers, and U-net architecture for segmentation tasks. While the U-net design creates a binary mask characterizing the tumor locations, the CNN architecture provides a binary output showing the presence or absence of tumors in a specific patch.

It is suggested to split the images into smaller patches and carry out the training procedure on these patches due to the computing difficulties involved in processing high-dimensional images in their entirety. On the basis of an extensive test dataset made up of 121,930 patches collected from five WSI pictures, experimental findings are given. The CNN model is assessed in a number of contexts, taking into account both cancerous and non-cancerous regions as well as partially cancerious regions. To evaluate the model's performance, metrics including average Intersection over Union (IoU), patch accuracy, patch F1 score, pixel accuracy, and pixel F1 score are produced.

The segmentation of tumor regions shows great accuracy at the picture level after results analysis. The CNN and U-net models are also explored, with emphasis on the benefits and drawbacks of each. This study sheds light on the use and assessment of CNN and U-net models for tumor area segmentation, highlighting their potential to provide accurate and effective tumurous site detection in medical applications.

Keywords: Tumor region segmentation, WSI(Whole slide imaging), U-net architecture, medical image processing, CNN (Convolutional Neural Network)

ÖZET

HİSTOPATOLOJİ GÖRÜNTÜLERINDE TÜMÖR BÖLGESININ BÖLÜTLENMESI

Emre ARSLANOĞLU
Muhanad TUAMEH

Bilgisayar Mühendisliği Bölümü
Bitirme Projesi

Danışman: Prof. Dr. Songül VARLI

Bu çalışma, raporda verilen WSI (tüm kesit görüntüsü) içindeki tümör bölgelerinin segmentasyonuna odaklanmaktadır ve bu modellerle elde edilen sonuçları sunmaktadır. Bu modellerden ilki, bir CNN katmanı ile oluşturulan ikili bir sınıflandırıcıdır ve diğeri segmentasyon için kullanılan U-net mimarisidir. CNN mimarisi, verilen yamalarda tümör olup olmadığını belirleyen ikili bir çıktı üretirken, U-net mimarisi tümör bölgelerini belirlerken ikili bir maske oluşturur.

Yüksek boyutlu görüntülerin tamamının aynı anda işlenmesiyle ilgili hesaplama zorlukları göz önüne alındığında, öncelikle görüntüleri küçük yamalara bölmek ve eğitimi bu yamalar üzerinde gerçekleştirmek önerilmektedir. Beş WSI görüntüsünden oluşan kapsamlı bir test veri setine dayanan deneysel sonuçlar sunulmaktadır ve bu veri setinde 121,930 yama bulunmaktadır. CNN modeli, tümör ve tümör olmayan yamaların yanı sıra kısmi tümör bölgelerinin dikkate alındığı farklı senaryolarda değerlendirilmektedir. Performans değerlendirme metrikleri olarak yama doğruluk, yama F1 skoru, piksel doğruluğu, piksel F1 skoru ve ortalama İzleme üzeri Birleşim (IoU) hesaplanarak modelin etkinliği değerlendirilir.

Sonuçların analizi sonucunda, tümör bölgelerinin segmentasyonu görüntü düzeyinde yüksek doğruluk göstermektedir. Hem U-net hem de CNN modelleri, avantajları ve dezavantajları vurgulanarak tartışılmıştır. Bu araştırma makalesi, tümör bölgesi segmentasyonu için CNN ve U-net modellerinin uygulanması ve değerlendirilmesine ilişkin değerli bilgiler sunarak, tıbbi uygulamalarda kanserli bölgelerin hassas ve

verimli tespiti için potansiyellerine vurgu yapmaktadır.

Anahtar Kelimeler: Tümör bölgesi segmentasyonu, Tüm kesit görüntüsü (WSI), U-net mimarisi, medikal görüntü işleme, evrişimli sinir ağları(CNN)

1
Introduction

Accurate and timely diagnosis of cancer is crucial to improve patient outcomes and facilitate effective treatment strategies [1]. In recent years, whole slide imaging (WSI) has emerged as a valuable tool in the fields of pathology, enabling the digital representation of entire tissue sections. Merging AI to pathology and computer-assisted analysis of tissue samples, will give WSI the potential to revolutionize cancer diagnostics and research [2].

The field of pathology has undergone a significant transformation with the arrival of digital pathology and the usage of Whole Slide Imaging (WSI) technology. Traditional pathology involved the examination of physical glass slides under a microscope, requiring pathologists to manually review and analyze individual slides. However, digital pathology has revolutionized this process by digitizing slides using WSI scanners, allowing for easier storage, sharing, and analysis of pathology images [3].

The transition from traditional pathology to digital pathology offers numerous advantages. First and foremost, WSI technology enables the creation of high-resolution digital images of entire slides, capturing every detail and cellular structure. These digital slides can be easily accessed and reviewed remotely, eliminating the need for physical transportation of glass slides and facilitating collaboration among pathologists and experts worldwide.

Analyzing Whole Slide Imaging (WSI) images presents distinct challenges primarily due to their high-resolution nature, which creates difficulties in reading and effectively using them for training AI models using traditional techniques such as Convolutional Neural Networks (CNN). Consequently, there is a growing need for developing robust and efficient computational methods to analyze and segment cancerous tissues from these images. Another limitation, is the cost of the equipment needed to analyze and train such data [3].

In this project, we aim to address this need by implementing and comparing

two approaches for tumor segmentation: pixel-wise segmentation and patch-wise classification. To this end, we will use state-of-the-art deep learning architectures, including U-net for pixel-wise segmentation and Convolutional Neural Network and ResNet for patch-wise classification. The effectiveness of these models will be evaluated using Intersection over Union (IoU) for pixel-wise segmentation and F-1 score for patch-wise classification. Through this comprehensive investigation, we hope to contribute to the ongoing efforts in cancer diagnostics and ultimately improve patient outcomes.

2
Preliminary

Tumors can be classified into two main types: benign tumors and malignant tumors. Benign tumors are typically localized and do not spread to other parts of the body, thus presenting minimal harm to the individual. On the other hand, malignant tumors, also known as cancerous tumors, have the ability to spread to other areas of the body, causing significant harm. In cancerous tumors, cells within the body grow abnormally and attack healthy tissues, leading to the formation of tumors [4].

Histopathological examination of tumors is a critical step in cancer diagnosis, involving the analysis of imaging data to identify the presence of tumor cells within healthy tissues. This process is typically performed manually and can be time-consuming, prone to errors, and subject to the pathologist's expertise. As such, automatic localization of tumor-rich areas is a critical component of computer-aided diagnostic systems [3]. Accurate identification of areas containing high densities of tumor cells can assist pathologists in assessing disease aggressiveness and selecting high-power areas for conditions such as tumor proliferation grading/scoring.

The liver is a visceral organ that is commonly affected by the metastatic spread of cancer. Early detection of liver cancer is crucial for effective management of the disease. Unfortunately, many people are unaware of their hepatitis status, which can increase their risk of developing liver cancer. Hepatocellular Carcinoma (HCC) is the most common form of primary liver cancer, accounting for approximately 90% of cases worldwide. HCC is a significant global health concern, and its incidence is on the rise both in Korea and globally. It is among the leading causes of cancer-related deaths worldwide, with the number of newly diagnosed cases increasing by 75% between 1990 and 2015. This increase can be attributed mainly to population growth and changes in age structures [5].

Because of the critical importance of accurately and rapidly detecting cancerous areas in liver cancer, numerous studies have been conducted in this area. Cancer segmentation models have been developed and published for various organs,

including the lung and pancreas [6, 7]. In this study, we will be using the PAIP2019 dataset, which was collected from individuals diagnosed with liver cancer between 2005 and 2018 at renowned institutions such as Seoul National University Hospital (SNUH), Seoul National University Bundang Hospital(SNUBH), and SMG-SNU Boramae Medical Center (BMC). The dataset includes images of resected tissue slides, which were shared as part of a challenge to facilitate research and development of effective solutions [5, 7]. The sharing of this dataset has spurred a wave of research, resulting in numerous solutions being proposed. In some studies, closed-source datasets collected by pathological research centers were also used.

Conducting research in the field of liver cancer segmentation presents several challenges. First, technical personnel must obtain legal permissions to collect data. Also, the high-resolution scanners used to transfer cancerous tissue images to digital media are exceedingly costly, and the image processing and machine learning techniques required for analyzing the digital media necessitate substantial storage space and processing power [7]. Open-source datasets have emerged as a viable solution for data collection and digital transfer. Furthermore, researchers have developed software patches and hardware solutions to meet the processing power requirements of the image processing and machine learning methods used in many studies. In some cases, visuals were divided into smaller parts for patching and analysis.

Another dataset used in this field is the Liver Tumor Segmentation (LiTS) dataset. This data set was created for lung cancer segmentation from CT images by universities and research centers such as Technical University of Munich (TUM), Radboud University Medical Center, Poly technique Montréal and CHUM Research Center, Sheba Medical Center, the Hebrew University of Jerusalem, Hadassah University Medical Center, IRCAD. It consists of a total of 201 CT images. While large lesion areas can be predicted with high accuracy in the studies performed, the success in small lesion areas is quite low. For this reason, Digital pathology images are to obtain more sensitive results than CT images [8–11].

In these study, two methods have been used U-Net and CNN to obtain results. During the evaluation phase, metrics such as Intersection over Union (IoU), Recall, Precision, and F-1 score were applied [8, 11]. The PAIP2019 data consists of two distinct tasks: segmentation of the total tumor region and segmentation of the viable tumor region. In the PAIP2019 challenge, while the evaluation method used for the first task is,

$$task\_1\_score = \begin{cases} jacard_i(GroundTruth_i, Prediction_i), & \text{if } jacard_i \geq T \\ 0, & \text{otherwise} \end{cases} \tag{2.1}$$

N denotes the number of cases,

$$task\_1\_aggregated = \frac{\sum_{i=0}^{N} task\_1\_score_i}{N} \tag{2.2}$$

the evaluation method used for the second task is

$$task\_2\_accuracy_i = 1.0 - \frac{(|groundtruth_i(\%) - prediction_i(\%)|)}{100.0} \tag{2.3}$$

$$task\_2\_weighted\_task\_score_i = task\_1\_score_i * task\_2\_accuracy_i \tag{2.4}$$

$$task\_2\_aggregated = \frac{\sum_{i=0}^{N} task\_2\_score_i}{N} \tag{2.5}$$

The maximum achieved score in the competition to date is 0.78 for the first task and 0.75 for the second task.[5]

In this study, the primary objective is to focus on the segmentation of tumor regions in liver cancer. The segmentation of the tumor regions can provide critical information to medical experts for the accurate diagnosis and treatment of liver cancer. Also, the results obtained from this study can be compared and evaluated by experts in the field or combined with other studies to improve the accuracy of the segmentation process. We believe that the results of this study can also serve as an intermediate step for the development of more comprehensive end-to-end projects in the field of liver cancer segmentation.

3
Feasibility

In this section, a feasibility study of the project was conducted, which aimed to determine the practicality and viability of the proposed endeavor. The study involved identifying the required resources and developing a time plan for the project's development. Other than that, a comprehensive examination of the legal and economic feasibility was conducted to ensure compliance with applicable regulations and financial sustainability.

3.1 Technical Feasibility

Regarding the technical feasibility, the software and hardware requirements of the project have been outlined to determine the practicality and feasibility of the proposed initiative. This includes an assessment of the necessary technical resources and infrastructure needed to successfully carry out the project.

3.1.1 Software Feasibility

In terms of software feasibility, the project will use the Python programming language because of its user-friendly nature, robust ecosystem, widespread use in academic research, and compatibility with the platforms on which the project will run. To implement the artificial neural networks required for the project, the PyTorch and Tensorflow libraries were chosen. Also, OpenCV, Pandas, and NumPy libraries were used in the patching and editing phase of large Whole Slide Images (WSI) used in the project.

3.1.2 Hardware Feasibility

Regarding hardware feasibility, the project demands high processing power to effectively train and test the Deep Learning models constructed for Computer Vision studies. This is a result of the extensive number of matrix operations required during

training and testing, as well as the need for repeated training phases to optimize hyperparameters.

Something else to note, the WSIs used in the project possess significantly higher memory requirements compared to traditional image processing projects, with individual images ranging from 600 MB to 2 GB. Therefore, it is necessary to have high memory and processing power to process images and use them in the training process. Despite certain software solutions being available for this problem, the demand for high processing power and memory persists.

Modern GPUs with high processing power are often used to satisfy these requirements. However, cloud services are preferred because of the high cost and power consumption of modern GPUs.

Among the available free platforms for training and testing the model, the Google cloud environment was selected. Google Colab was preferred as it seamlessly integrates with the Google Drive environment and offers processing power and memory capabilities. The specification of the machine provided by Google Colab are listed in Table 3.1.

GPU	Tesla K80 GPU
CPU	Intel(R) Xeon(R) CPU @ 2.20GHz
RAM	12.7 GB
Time	12 Hours per Week
Storage	256 GB

Table 3.1 System Features on the Colab Platform

If a cloud environment is not used, the following system requirements are recommended in Table 3.2.

GPU	NVIDIA RTX 4090
CPU	Intel(R) i9 13th
RAM	32 GB
Storage	256 GB

Table 3.2 Recommended System Requirements

In order to speed up the training process an instance of Google Cloud Platform (GCP) was created using the free tier GCP provides for the new users. The specification of the used instance is shown in 3.3

Machine type	g2-standard-8 (8 vCPU, 32 GB memory)
CPU platform	Intel Cascade Lake
GPU	NVIDIA L4
Operating system	Ubuntu 23.04
Storage	500 GB

Table 3.3 Specification of the Created Instance

3.2 Time Feasibility

The project's requirements and the project's timeline were discussed beginning at the end of February. In March, meetings with the advisor were held to go over the details of the project. Once the project's subject and specifics were established, the requisite feasibility studies were conducted, and problems found during the hardware and software feasibility assessments were addressed. Also, necessary information was gathered through study and training about the platforms, programming languages, and frameworks that will be used.

Something else to note, a request for the dataset needed for the project was made to the data source. After obtaining the required access from the data provider, the preliminary review process was successfully finished, and the dataset was obtained. Hyperparameter optimization, model comparisons, and training trials were then scheduled after carrying out the essential preparation processes for data processing, model building, and training.

The project is anticipated to be finished in nine weeks after the dataset is received. The Figure 3.1 provides the process's Gantt diagram.

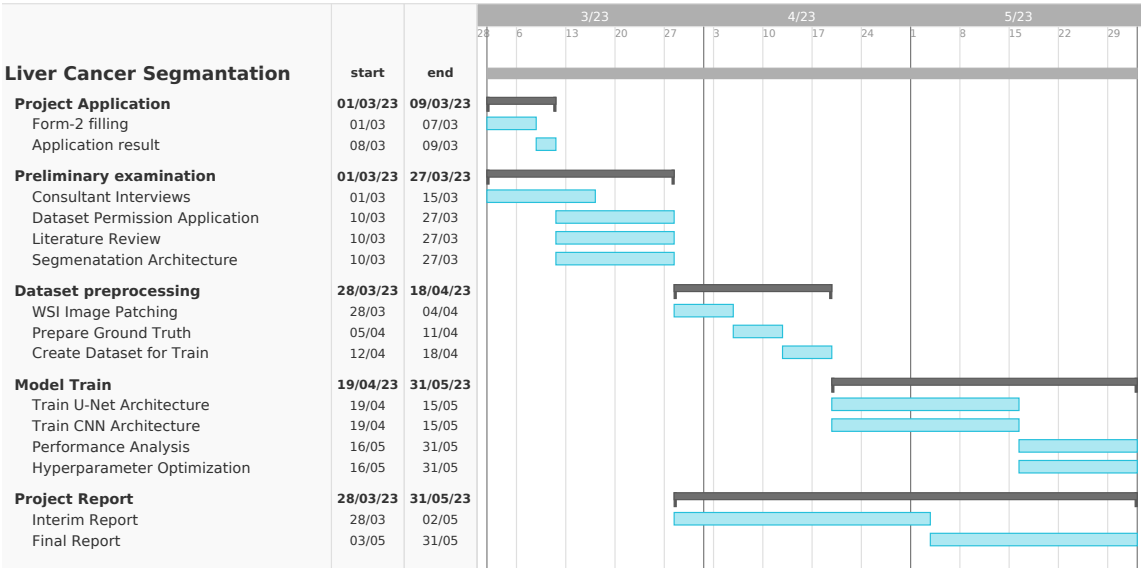


Figure 3.1 Gantt Diagram of the Project

3.3 Legal Feasibility

In terms of legal feasibility, it should be noted that the dataset used for the project contains images of 50 resected cells from individuals diagnosed with hepatocellular carcinoma of the liver. These cell images were scanned using digital pathology scanners and transferred to the computer environment. All cases represent tumor tissues of the liver diagnosed at SNUH, SNUBH, and SMG-SNU BMC from 2005 to June 2018. To protect patients' privacy, all personal labels of scan images were removed.

It is important to note that the collected PAIP2019 dataset was made available for competition on the grand challenge site. Access was provided to individuals who are interested in working in this field. It is crucial to comply with the regulations regarding the use of medical data and to obtain the necessary permissions and approvals before using any medical data for research purposes. Therefore, the necessary legal requirements should be carefully considered and met throughout the project's development process to ensure compliance with relevant regulations and laws.[5]

3.4 Economical Feasibility:

The dataset used in the project is a publicly available dataset. In addition, since the software used is open source and no other licensed software is used, there is no cost in the software and dataset stage.

In the hardware phase, the training will be carried out in the cloud environment. However, because of the large size of the dataset, it was foreseen that the features offered free of charge during the storage and training phases would not be sufficient. For this reason, extra memory and storage services will be obtained from the preferred cloud platform. Although these are not high-paid services, they can appear as an expense between 100-150 TL per month. However, it is worth noting that the cost estimation provided in Turkish Lira (TL) may vary depending on the exchange rate fluctuations and the preferred cloud platform's pricing policies.

4
System Analysis

This section provides a general overview of the project and defines its objectives and requirements. Additionally, the performance metrics that will be applied to assess the success of the models are outlined. Furthermore, a comprehensive analysis of the dataset's content and structure is conducted to determine the optimal architecture for addressing the task at hand.

4.1 Requirements Analysis

The primary goal of this project is to develop a reliable and efficient method for segmenting cancerous tissues from whole slide images (WSI). WSI refers to high-resolution digital images of entire tissue sections from glass slides that are typically viewed under a microscope. These digital images can be viewed and analyzed on a computer, allowing for remote consultation, telepathology, and computer-assisted analysis of the tissue samples. Whole slide images are commonly used in fields such as pathology, histology, and cytology for diagnosis, research, and education purposes.

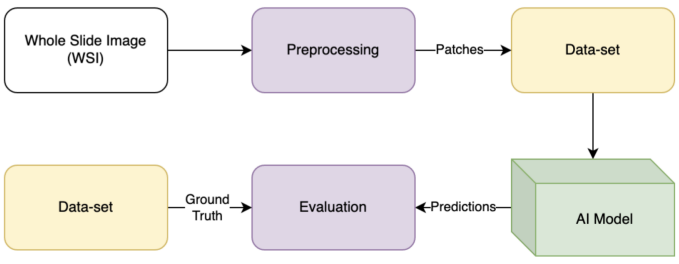


Figure 4.1 Block Schema of The Proposed Pipeline

Given the high-resolution nature of WSI images, they are typically too large to be directly input into a model. Consequently, the proposed pipeline is as shown in Figure 4.1, WSIs are divided into smaller images (256x256 pixels) called patches, along with corresponding mask images that indicate the viable tumor area. These mask patches serve as the ground truth for tumor regions. The generated patches are then given to an AI Model to be trained. The AI architecture used in this project can be U-net for

pixel-wise classification, CNN and ResNet for patch-wise classification. The results of both models are then evaluated using several metrics such as IoU and F1-score.

4.2 Dataset Analysis

This section presents the results obtained from the analysis of the model's dataset. It should be noted that the WSIs in the dataset varies in sizes, resulting in a different number of patches for each WSI. Following the removal of background regions, the average number of patches per image in the training dataset is found to be 20,968, with a minimum of 8,894 patches and a maximum of 38,365 patches. The total number of patches in the train and validation sets, after the identification and removal of background regions, is 943,596. Among these patches, 286,838 correspond to patches with over 80% tumor area, while 590,882 patches consist entirely of tumor-free regions. The remaining 65,876 patches contain less than 80% tumor region. In the CNN architecture, 656,758 patches with tumor regions above the threshold or completely tumor-free are used, whereas the U-net architecture uses all 943,596 images. The distribution of these patches is visualized in Figure 4.2.

Upon examining the patch distribution, it becomes apparent that the dataset contains an average of 30% patches with cancerous areas. This proportion ranges from a maximum of 66% and to a minimum of cancerious area 2%. These findings indicate an imbalance in the dataset, which can affect the model's ability to accurately detect cancerous regions. To address this issue and enhance the detection performance of cancerous regions, techniques such as data augmentation or sampling methods are used, aiming to create a more balanced dataset.

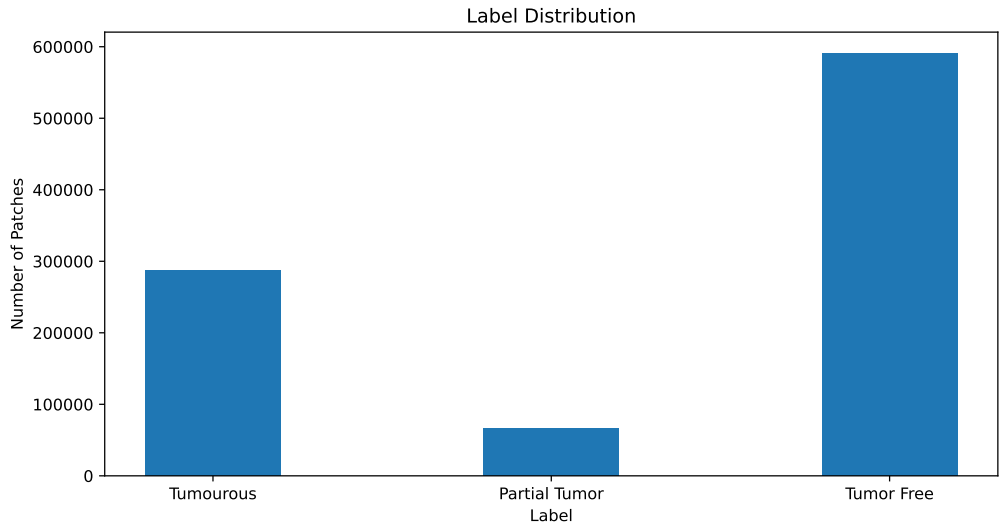


Figure 4.2 Cancer Area Rate

The patching process results can be observed in Figure 4.3.a and 4.3.b. Figure 4.3.a displays patches containing cancerous tissues with at least 80% cancerous area, while Figure 4.3.b shows the corresponding mask of each patch, highlighting the cancerous regions. In contrast, Figure 4.3.c showcases patches of healthy tissues without significant cancerous areas.

4.3 Architecture Analysis

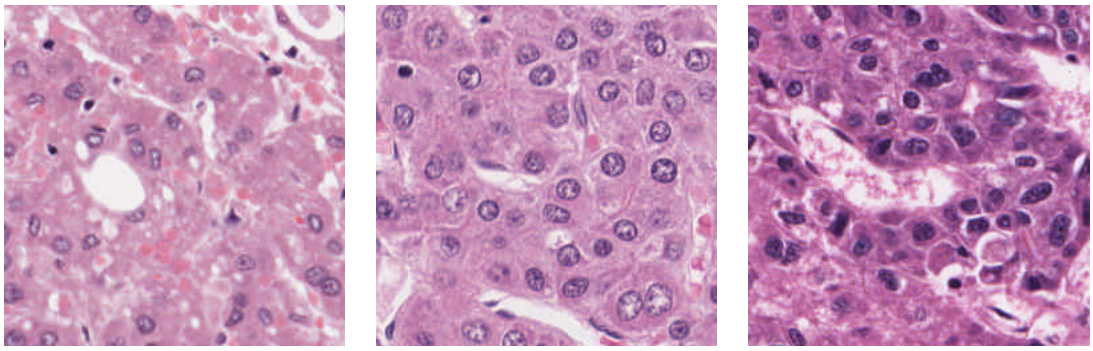
Two methods can be considered for segmenting the tumor area: pixel-wise segmentation and patch-wise classification. Pixel-wise segmentation involves classifying each pixel within a patch as either cancerous or non-cancerous. Although this approach tends to result in more accurate results, the computational cost of training the model can be huge because of the large number of pixels processed. Alternatively, patch-wise classification classify entire patches as cancerous or non-cancerous. While this method may produce less precise tumor segmentation, it is more computationally efficient as it processes fewer data points.

In this project, both pixel-wise segmentation and patch-wise classification methods are implemented, and their results are compared. The U-net architecture is used for pixel-wise segmentation because of its advantages in medical image segmentation tasks, particularly when segmenting cancerous tissues. Whereas, the CNN and ResNet architecture is used for patch-wise classification.

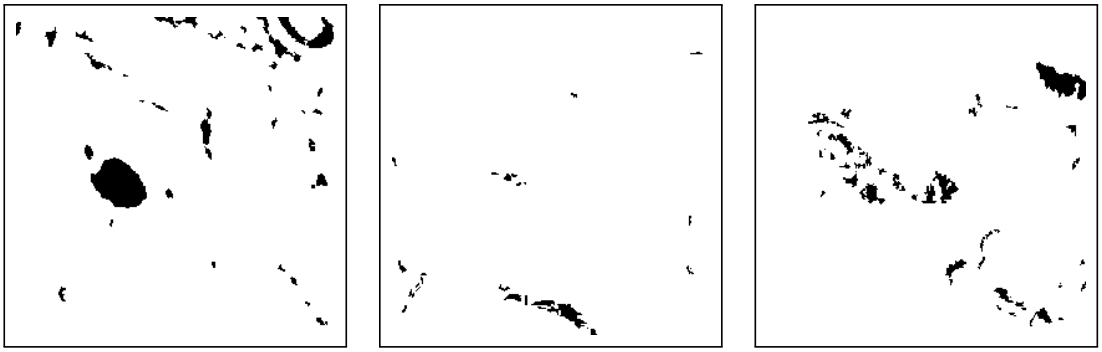
The outcomes of the two segmentation methods are different in nature. For pixel-wise segmentation, the output is a mask image resembling the patches illustrated in Figure 4.3.a, which highlights the cancerous regions within the patch. In this approach, the model identifies and classifies each pixel as either cancerous or non-cancerous, providing a detailed representation of the tumor area.

On the other hand, patch-wise classification results in a binary classification, where the output is either 1 (if the patch contains 80% or more cancerous area) or 0 (if the patch not contains any tumorous pixel.).

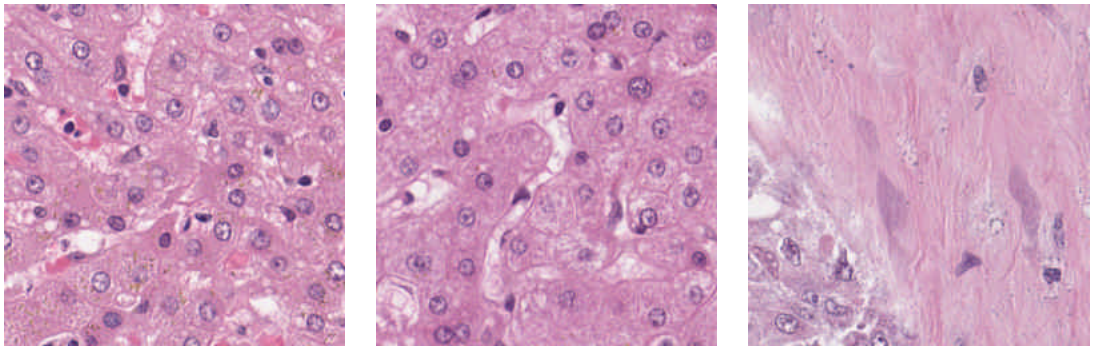
Upon obtaining the segmentation results, the patches are concatenated to reconstruct the whole slide image with annotated cancerous regions. This reconstructed image allows for a comprehensive visualization of the cancerous areas within the tissue sample, providing valuable insights for diagnostic and research purposes. By comparing the results of both segmentation methods, the most suitable approach for the specific application can be determined, balancing the trade-offs between precision and computational efficiency.



(a) Contains 80% cancerous area



(b) Mask of Cancerous Area



(c) 80% Free of Cancerous Areas

Figure 4.3 Patched Image Examples

The performance of the models is assessed using Intersection over Union (IoU) for pixel-wise segmentation and F-1 score for patch-wise classification. IoU is a metric that measures the overlap between the predicted bounding box and the ground truth bounding box of an object, while the F-1 score is a measure of a model's accuracy, taking into account both precision and recall.

5
System design

5.1 Dataset Design

The dataset used in this project is derived from the publicly available PAIP2019 challenge dataset. It built of 50 whole slide images (WSIs) for training, 10 for validation, and 20 for testing. The WSIs in this dataset represent tumor tissue samples from patients who were histologically diagnosed with hepatocellular carcinoma of the liver at several hospitals in South Korea between 2005 and 2018.

The dataset provides original whole slide image, XML annotation made by pathologists, ground-truth binary pixel masks, which are generated from the XML annotation, for whole tumor area and viable tumor area, and viable tumor burden calculated from the binary pixel masks. The dataset provides the following components:

- Original whole slide image (WSI): High-resolution scanned images compressed in the SVS format. In addition, each picture has a size of approximately 43000x45000 pixels.
- XML annotations: Annotations made by pathologists, highlighting the tumor regions.
- Ground-truth binary pixel masks: Generated from the XML annotations, these masks correspond to the whole tumor area and the viable tumor area.
- Viable tumor burden: Calculated from the binary pixel masks, this metric represents the area of the whole tumor, the sum of the viable tumor area, and the ratio of the sum of the viable tumor area to the whole tumor area.

Figure 5.1 illustrates a complete image, wherein the yellow line outlines the tumor region and the green line delineates the cancerous tissues. The objective of the project is to accurately segment the regions marked with the green line. The dataset consists

of 50 complete images, with an average size of 43000x45000 pixels. Because of their large size, these images cannot be directly used in models, and as such, the dataset undergoes several preprocessing steps to be used in the model.

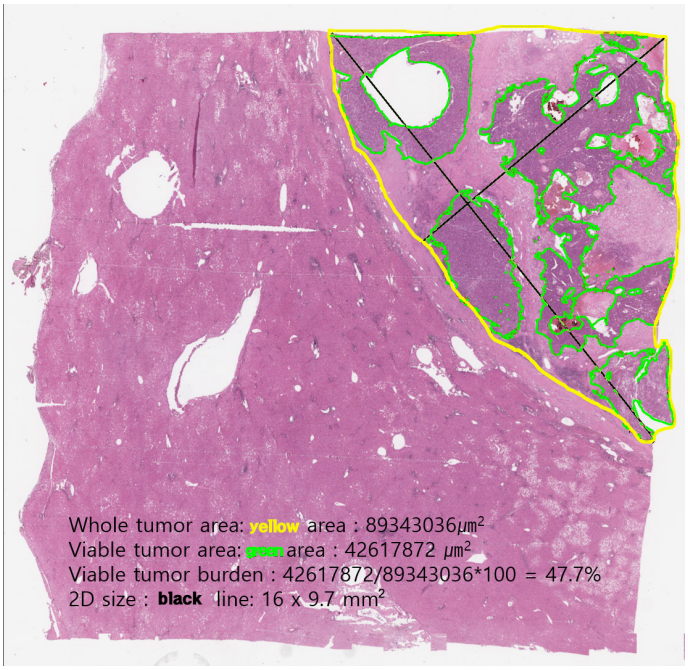


Figure 5.1 WSI Image

5.2 Preprocessing Data

The whole-slide images (WSI) in this study are divided into small patches, each containing a 256x256 image and its corresponding pixel-based label mask. The size of these patches is a hyperparameter that directly impacts the model's performance. The WSIs and masks are systematically patched and saved to the disk to be used during the training and testing phases. An important step during the patching process involves determining whether a fragment belongs to the cell or background. To this end, if the average pixel value in all three channels of an image exceeds 225, it is classified as the background and not saved to disk. Background areas do not provide valuable information for training the model. The recorded images are processed and used separately in accordance with the U-net and CNN architectures to be developed.

To prepare the dataset for the CNN architecture, the labels of the patches need to be specified. Some patches correspond to tumor-free regions, while others may contain partially tumorous areas. Consequently, a threshold is established to identify tumor regions. If a patch contains 80% or more tumor cells, it is labeled as a tumor patch. In the CNN model, only patches with over 80% tumor or no tumor regions are used. By using these patches, the model is expected to learn to identify tumor regions. on

the other hand, in the U-net architecture, different distributions of cancerous and non-cancerous patches are used to find out the best distribution. Unlike the CNN architecture, the U-net does not require explicit labeling as the patched mask images serve as ground truth data. During the dataset creation stage, labeling and filtering procedures are used to process the patch statuses.

The created dataset consists of 943,596 images. Although this number decreases as the dataset is filtered for specific architectures, it still contains a substantial amount of data for memory storage. Therefore, direct storage of patches is impractical. As a solution, a CSV file containing the addresses and label information of the dataset is generated. During the training phase, the images that need to be stored in memory are read and provided to the model using this CSV file. The CSV document includes the following particulars in sequential order.

After the patching process, the following dataset is constructed for model training:

- Image: The name of patch image.
- Row: The horizontal order of the patch in the original WSI.
- Col: The vertical order of the patch in the original WSI.
- Cancer: A binary label indicating if the patch is cancerous (1) or non-cancerous (0), if the patch is partially tumorous(Below 80%) (-1)
- Path: The file path used for reading the patch.

5.3 Handling the Imbalance in the Training Dataset

In the training dataset, there is a noticeable imbalance between the number of noncancerous patches and cancerous ones. This imbalance may adversely affect the training process, as the model may become biased towards the majority class (noncancerous patches) and might not perform well in identifying the minority class (cancerous patches). This issue could lead to a higher false-negative rate, which is not desirable in medical applications where accurate detection of cancerous regions is crucial.

To mitigate the effects of this class imbalance, several techniques can be used for both pixel-wise segmentation and patch-wise classification tasks:

Data augmentation: For both pixel-wise segmentation and patch-wise classification, data augmentation can be applied to artificially increase the number of cancerous

patches in the dataset. This can be done by applying various transformations, such as rotation, scaling, and flipping, to the existing cancerous patches. This technique helps to balance the class distribution and provides more diverse training samples for the model to learn from.

Undersampling and oversampling: For patch-wise classification, undersampling and oversampling techniques can be used to balance the class distribution. Undersampling involves removing some noncancerous patches from the dataset, while oversampling includes duplicating some of the cancerous patches. However, these methods should be used cautiously, as undersampling may lead to loss of valuable information and oversampling might result in overfitting.

5.4 Software Design

In this project, two primary deep learning architectures are used: U-net for pixel-wise segmentation and a standard Convolutional Neural Network (CNN) for patch-wise classification.

5.4.1 U-net Architecture

U-net is a widely-used and well-established convolutional neural network (CNN) architecture specifically designed for image segmentation tasks. It has demonstrated exceptional performance in medical image segmentation applications, making it an ideal choice for this project.

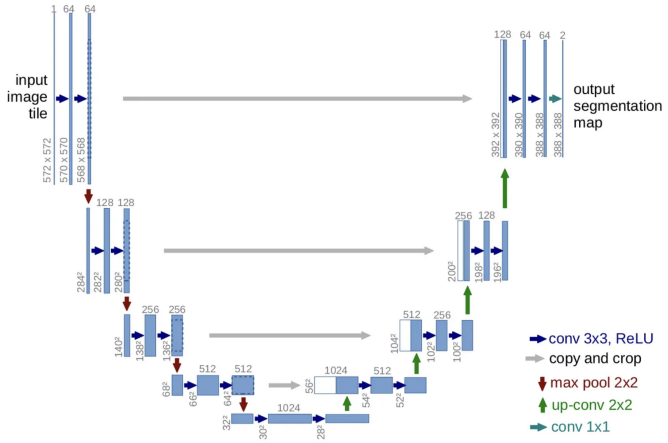


Figure 5.2 U-net architecture [12]

The U-net architecture has two principal components: the contraction path (encoder) and the expansion path (decoder). The encoder consists of multiple convolutional and max-pooling layers that progressively reduce the spatial dimensions of the input

image while increasing the number of feature maps. This enables the model to capture more high-level abstract features from the image [12].

In contrast, the decoder contains several up-convolutional layers that incrementally increase the spatial dimensions of the feature maps while decreasing their count. To retain more detailed information for segmentation, feature maps from corresponding layers in the encoder are concatenated with the up-convolutional layers in the decoder.

U-net's skip connections are particularly beneficial for image segmentation tasks, as they allow the network to capture both local and global features of the image. By passing learned features from the encoder to the corresponding decoder layer, the skip connections help preserve spatial information lost during the down-sampling process.

Max Pooling: Max pooling is the process of representing the values within a selected NxN region of the activation map with the maximum value in that region. The goal is to accelerate the training process and prevent overfitting by representing regions with their most dominant values.

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 6 & 8 \\ 3 & 6 & 9 & 12 \\ 4 & 8 & 12 & 16 \end{bmatrix} \rightarrow_{\max} \begin{bmatrix} 4 & 8 \\ 8 & 16 \end{bmatrix}$$

Figure 5.3 Max pooling

Convolution: Convolution is a mathematical operation that is often used in deep learning for image and signal processing. It involves sliding a small matrix, called a kernel or filter, over an input image or signal and computing a dot product between the kernel and the input values within the window of the kernel. This process results in a feature map that highlights specific patterns or features within the input. Convolutional neural networks (CNNs) uses this operation in their layers to extract meaningful features from images and other types of data.

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 6 \\ 4 & 5 & 6 & 7 \end{bmatrix} * \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 6 & 8 \\ 6 & 8 & 10 \\ 8 & 10 & 12 \end{bmatrix}$$

Figure 5.4 Convulution

Upsampling: Up-convolution layers increase the spatial resolution of the feature maps, allowing the network to produce a segmentation mask with dimensions similar to

the input image. This upsampling process is achieved by applying a transposed convolution operation that expands the feature map's spatial dimensions.

Skip connections: Up-convolution layers are combined with skip connections from the corresponding layers in the encoder. These connections provide the up-convolution layers with additional spatial information that was lost during the down-sampling process in the encoder. By concatenating these feature maps, the decoder can generate a more accurate and detailed segmentation mask.

For this project, U-net is an appropriate choice due to its ability to accurately segment cancerous and non-cancerous regions in tissue patches. The architecture's capacity to capture both local and global features and preserve spatial information renders it highly suitable for this task.

5.4.2 Convolutional neural network (CNN)

Convolutional Neural Networks (CNNs) are a class of neural networks specifically designed for image processing tasks, such as image classification. By leveraging the spatial correlation between adjacent pixels, CNNs are capable of handling high-dimensional inputs, such as images.

In the context of patch-wise classification for cancer tissues, CNNs are an excellent choice because they can effectively learn features from local regions within the image, which is crucial for accurate patch classification. As cancerous tissue can exhibit various visual appearances and patterns, CNNs can exploit the spatial correlation between adjacent pixels to learn more robust features. Furthermore, by processing patches in parallel, CNNs can efficiently handle large volumes of data.

A typical CNN consists of several layers, including convolutional layers, activation functions, pooling layers, and fully connected layers. Convolutional layers are designed to learn local features from the input image by applying a set of filters or kernels across the image. Activation functions introduce non-linearity into the network, enabling it to learn complex patterns. Pooling layers are used to reduce the spatial dimensions of the feature maps, while fully connected layers are responsible for the final classification decision.

The final stage of convolutional neural networks involves classifying images using the distinctive features extracted in the convolutional layers. This process is performed using linear layers, also known as fully-connected or dense layers. The reason for this is that each neuron in these layers is connected to all the neurons in the previous and next layers [13].

The performance of the CNN can be optimized by fine-tuning various hyperparameters, such as the number of layers, number of filters, filter size, activation function, and learning rate. By adjusting these parameters, the model can be tailored to achieve the best possible performance on the given task.

6
Implementation

The project is being modeled and trained using two distinct methods. One of these models is a segmentation model using the U-Net architecture, while the other model incorporates deep learning layers, specifically Convolutional Neural Network (CNN) layers.

6.1 CNN Architecture

In the context of the current research project concerning tumor area segmentation, the deep neural network model has been formulated using the convolutional neural network (CNN) architecture. Patches of dimensions 256x256 pixels serve as the input for this model. The principal objective of this binary classifier architecture is to determine whether a given patch corresponds to the cancerous region. The effectiveness of the model is assessed through various metrics, namely patch-based accuracy, patch-based F1 score, pixel-based accuracy, pixel-based F1 score, and the Intersection over Union (IoU) metric. The success of the model is contingent upon multiple hyperparameters, including the size of the pixels used at this stage. Table 6.1 provides a comprehensive overview of the parameters used within the model.

Table 6.1 Hyperparameters

Train / Test Split	%90-%10
Epoch	Adaptif
Optimizer	Adam
Batch Size	32
Criterion	BCE Loss
Patch Size	256x256

The architecture itself has four CNN layers, with 32, 32, 64, and 64 3x3 filters, respectively. The intermediate layers use the Rectified Linear Unit (ReLU) function, while the output layer uses the Sigmoid function. In order to mitigate issues associated with overfitting, batch normalization and dropout techniques are implemented at the

output of each layer. During the training of the model, a dropout rate of 0.2 is adopted. The architectural diagram is depicted in Figure 6.2.

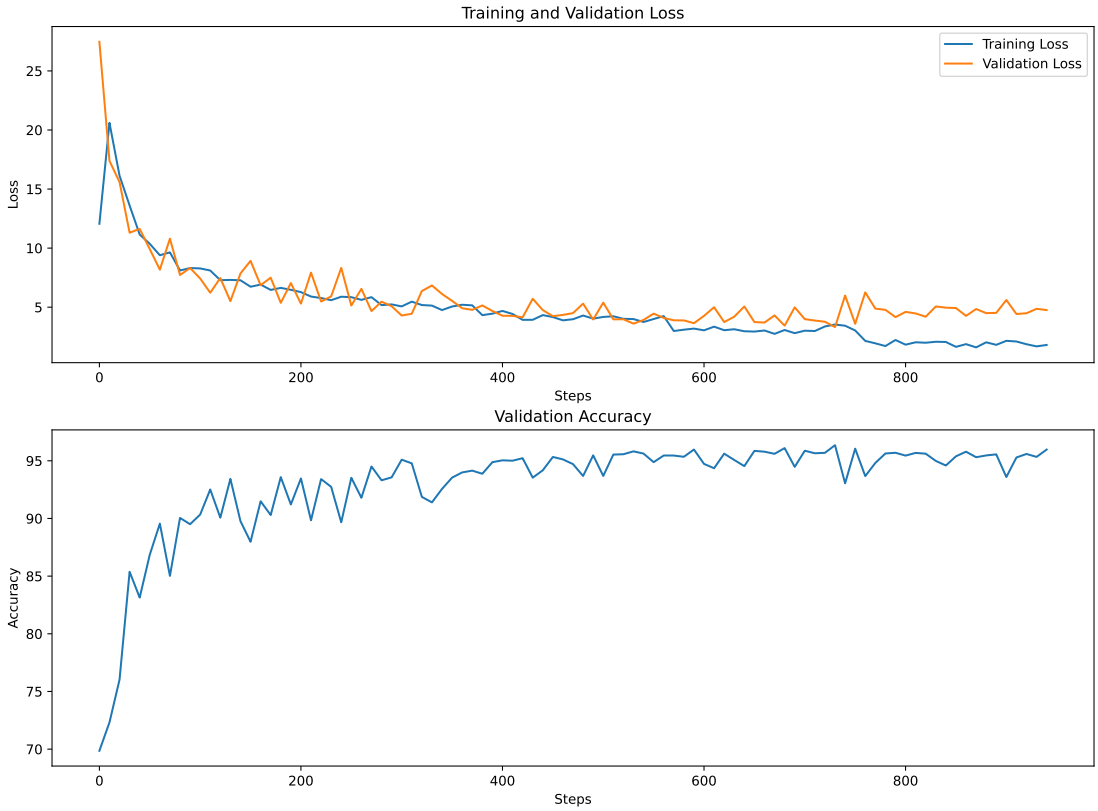


Figure 6.1 Loss and Accuracy of CNN

The primary challenge encountered during the coding process was the large size of the image data, rendering it impractical to store all the images in memory. To overcome this obstacle, the implementation phase contains a dataset consisting of image paths and corresponding labels, rather than the actual images. During training, images are read from the provided paths in batches, thereby reducing the memory requirements. Model and optimizer information is saved after each epoch, and the results are validated using a predetermined validation dataset containing 300 batches. After that, the testing phase is conducted using designated test images and a dataset containing the corresponding paths and label information. Similar to the training phase, test images are read in batches based on the previously recorded model. The obtained results are evaluated by using various metrics at both the patch and pixel levels.

In the patch-based experiments, patches containing 80% or more cancerous regions, as identified during the label removal process, are labeled as cancerous, while sections

without any cancer-free regions are labeled accordingly. It should be noted that certain patches may contain less than 80% cancerous areas. These patches present a challenge during the testing phase as they introduce a margin of error in the patch-based classification. In some cases, the model may incorrectly classify cancerous regions as non-cancerous or vice versa if it correctly predicts the label. Consequently, the patch-based evaluations include an inherent margin of error.

To assess the model's performance, metrics such as accuracy and F1 score are used at the patch level. At the pixel level, accuracy and F1 score are calculated by generating masks of size 256x256 using the estimated labels. Furthermore, the IoU metric is computed based on the generated masks. The resulting metrics are presented in the Results section of the report.

Furthermore, Figure 6.1 depicts the train and validation loss calculation of the model during training, along with the validation accuracy values. The graph illustrates that the loss decreases simultaneously for both the validation and train data. This observation indicates that the model is not overfitting, as the performance on the validation set aligns with the training set.

By altering the ResNet18 model's fully connected layer after it was developed using the ImageNet dataset, we performed transfer learning in conjunction with our CNN model. We used the pre-trained and freezed CNN layers of the ResNet 18 architecture in our model. Additionally, we added a three-layer structure to the fully connected layer and trained the architecture's classifier component on our dataset.

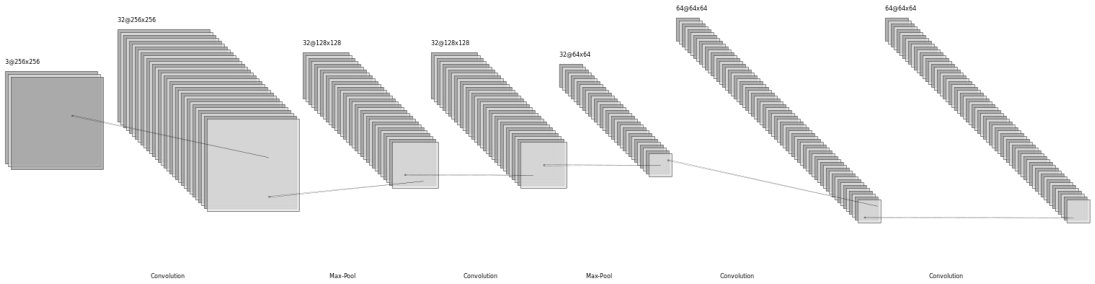


Figure 6.2 Convolutional Neural Network Architecture

6.2 U-net Architecture

Besides, the U-net architecture, despite sharing the same input as the CNN, return a distinct output. Instead of providing a binary value like the CNN, the U-net generates a binary mask with dimensions identical to the input patch, segmenting the cancerous region. The performance of the U-net can be evaluated using (IoU) score, comparing its output with the ground-truth binary mask.

Because of the huge size of the dataset, training the model in a single session becomes infeasible, primarily because of the limitations of cloud computing runtime restrictions. Cloud computing platforms typically enforce constraints on the maximum runtime for a single session, which may not be sufficient for the extensive training required for large datasets. To circumvent this issue, an effective strategy involves saving the model at specific intervals, typically after a predetermined number of epochs. By doing so, the model's current state is preserved, allowing the training process to resume from the last saved checkpoint in next sessions. This approach ensures that the model can be trained on the large dataset Despite the constraints of cloud computing platforms, leading to improved performance and accuracy in cancerous tissue segmentation tasks.

Later in the project GCP was used for training U-Net model on the whole data. GCP provided us with a good GPU that can train the model fast and with no constraints.

7
Experimental Results

In this section of the report, the results of two different models trained on the test dataset consisting of 5 whole-slide images (WSI) are presented. The test dataset, after patching and background removal, contains a total of 121,930 patches. Among these patches, 28,352 contain tumors, 75,551 are completely tumor-free, and 18,027 are partially tumorous as shown in 7.1.

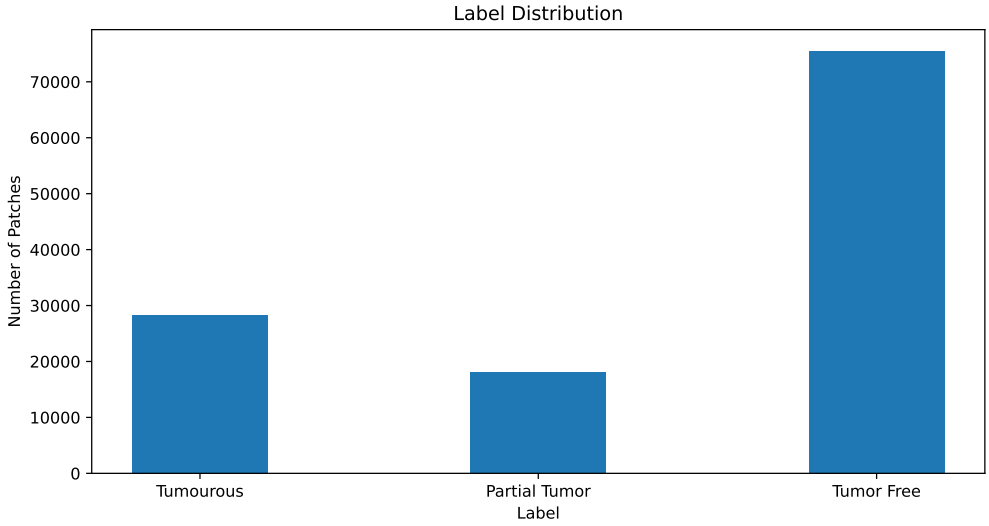


Figure 7.1 Cancer Area Rate of Test Dataset

7.1 Evaluation of CNN Architecture Performance

The experimental outcomes of the CNN model are documented across three distinct sets of model results. The initial set pertains to the CNN model trained on a dataset comprising 20 Whole Slide Imaging (WSI) images. When evaluating the test data with the consideration that partially tumorous areas are deemed tumor-free, the corresponding results are presented in Table 7.2. Conversely, when including partially affected tumor areas within the tumor area, the outcomes are provided in Table 7.1.

The second model is a CNN network that trained on a dataset consisting of 45 WSI images. The outcomes for this model can be observed in Tables 7.3 and 7.4

Lastly, the final model incorporates a pre-trained ResNet18 model that was initially trained with 45 WSI images. Subsequently, the fully connected layer of the ResNet18 model was fine-tuned. The results obtained from this model can be found in Tables 7.5 and 7.5.

The reported results have been obtained using two distinct approaches: path-based and pixel-based. The evaluation metrics employed include accuracy and F1 score, both on a pixel and patch basis. Additionally, the Intersection over Union (IoU) metric has been calculated to assess the performance in identifying cancerous regions.

Image Name	Patch Accuracy	Patch F1 score	Pixel Accuracy	Pixel F1 score	IoU
Training_phase_2_046	77.22	78.92	74.54	70.65	66.11
Training_phase_2_047	63.45	31.28	71.49	62.74	62.20
Training_phase_2_048	94.69	90.05	94.74	94.21	93.87
Training_phase_2_049	88.43	83.01	87.55	86.27	84.67
Training_phase_2_050	82.28	63.44	82.31	81.10	80.32

Table 7.1 The Model Trained with 20 Images and Partially Tumorous Areas Labelled as Tumorous Areas

Image Name	Patch Accuracy	Patch F1 score	Pixel Accuracy	Pixel F1 score	IoU
Training_phase_2_046	64.23	40.75	74.54	70.65	66.11
Training_phase_2_047	74.89	34.74	71.49	62.74	62.20
Training_phase_2_048	94.79	90.07	94.74	94.21	93.87
Training_phase_2_049	88.79	81.36	87.55	86.27	84.67
Training_phase_2_050	82.85	60.58	82.31	81.10	80.32

Table 7.2 The Model Trained with 20 Images and Partially Tumorous Areas Labelled as Non-Tumorous

Image Name	Patch Accuracy	Patch F1 score	Pixel Accuracy	Pixel F1 score	IoU
Training_phase_2_046	91.68	92.94	81.31	82.77	76.72
Training_phase_2_047	62.16	24.49	70.50	61.52	61.06
Training_phase_2_048	98.67	97.30	98.44	98.09	97.72
Training_phase_2_049	89.51	85.17	86.73	86.63	84.79
Training_phase_2_050	87.78	67.76	88.27	86.74	86.05

Table 7.3 The Model Trained with 45 Images and Partially Tumorous Areas Labelled as Tumorous

7.2 Evaluation of U-Net Performance

The original plan was to only train the U-Net model on patches with malignant regions. Given that these patches include both cancerous and non-cancerous parts, the underlying assumption was that such a technique would naturally produce balanced

Image Name	Patch Accuracy	Patch F1 score	Pixel Accuracy	Pixel F1 score	IoU
Training_phase_2_046	56.96	39.28	81.31	82.77	76.72
Training_phase_2_047	73.53	26.15	70.50	61.52	61.06
Training_phase_2_048	98.27	96.33	98.44	98.09	97.72
Training_phase_2_049	86.72	79.22	86.73	86.63	84.79
Training_phase_2_050	88.82	66.77	88.27	86.74	86.05

Table 7.4 The Model Trained with 45 Images and Partially Tumorous Areas Labelled as Non-Tumorous

Image Name	Patch Accuracy	Patch F1 score	Pixel Accuracy	Pixel F1 score	IoU
Training_phase_2_046	82.59	83.91	78.16	75.53	70.60
Training_phase_2_047	90.37	88.01	87.60	86.25	83.71
Training_phase_2_048	97.76	95.35	98.18	97.41	97.13
Training_phase_2_049	91.72	86.17	92.11	89.97	88.59
Training_phase_2_050	91.09	70.62	92.73	90.43	89.93

Table 7.5 The Resnet Model Trained with 45 Images and Partially Tumorous Areas Labelled as Tumorous

Image Name	Patch Accuracy	Patch F1 score	Pixel Accuracy	Pixel F1 score	IoU
Training_phase_2_046	63.84	41.36	78.16	75.53	70.60
Training_phase_2_047	86.99	80.19	87.60	86.25	83.71
Training_phase_2_048	98.58	96.92	98.18	97.41	97.13
Training_phase_2_049	94.32	89.22	92.11	89.97	88.59
Training_phase_2_050	94.33	77.18	92.73	90.43	89.93

Table 7.6 The Resnet Model Trained with 45 Images and Partially Tumorous Areas Labelled as Non-Tumorous

data, since that these patches contain both cancerous and non-cancerous regions. However, after some experiments a bias in favor of the malignant data was produced by training the model just on cancerous patches.

As shown in Table 7.7, the model's capacity to distinguish between benign patches and non-cancerous pixels was negatively impacted by the training bias.It is essential to have a balanced training dataset that includes both malignant and benign patches in order to stay away from bias.

The next question that came up was how many benign patches should be included in the training process. This is important because, it's critical to prevent an excessive bias towards the cancerous patches, but it's also necessary to prevent shifting the scales extremely in favor of benign data. As a result, it is decided to run several experiments by training the model on diverse data distributions and find out the best fit.

Also, to measure the impact of dataset size on the model's performance, experiments were conducted using two distinct volumes of data. The U-Net model was trained on two datasets, one using 20 Whole Slide Images (WSIs) and another U-Net using 45 WSIs, each contains different data distributions. This experiment aimed to evaluate

the influence of increased data volume on model performance, providing insights into a better data distribution for effective model training.

Table 7.7 the results of the model trained on 20 WSIs independently for each WSI. Whereas, Table 7.8 shows the overall results.

WSI Number	Trained Data	Cancerous Area					Benign Area					Overall				
		Mean IoU	Accuracy	F1 Score	Recall	Precision	Mean IoU	Accuracy	F1 Score	Recall	Precision	Mean IoU	Accuracy	F1 Score	Recall	Precision
46	50% Cancerous 50% benign	60.5%	77.3%	78.3%	77.3%	82.0%	39.5%	79.0%	87.1%	79.0%	100.0%	53.0%	77.9%	81.4%	77.9%	88.4%
	67% Cancerous 33% benign	58.3%	76.1%	76.3%	76.8%	86.7%	33.7%	67.4%	78.6%	100.0%	2.5%	49.5%	73.0%	77.1%	85.0%	56.6%
	100% Cancerous	34.7%	69.5%	57.5%	100.0%	69.5%	0%	0%	0%	100.0%	0%	22.3%	44.7%	37.0%	100.0%	44.7%
47	50% Cancerous 50% benign	31.3%	50.8%	45.4%	50.8%	76.4%	49.5%%	99.0%	99.5%	99.0%	100.0%	37.8%	68.0%	64.7%	68.0%	84.8%
	67% Cancerous 33% benign	26.0%	45.6%	36.9%	10.1%	85.4%	49.5%	99.1%	99.5%	100.0%	4.2%	34.4%	64.7%	59.2%	42.2%	56.4%
	100% Cancerous	29.5%	59.1%	45.6%	75.6%	59.1%	0%	0%	0%	100.0%	0%	18.9%	38.0%	29.3%	84.3%	38.0%
48	50% Cancerous 50% benign	40.4%	60.2%	60.3%	60.2%	80.1%	49.4%	98.8%	99.3%	98.8%	100.0%	43.6%	74.0%	74.2%	74.0%	87.2%
	67% Cancerous 33% benign	50.4%	70.6%	72.6%	53.1%	85.3%	48.9%	97.9%	98.9%	100.0%	1.8%	49.9%	80.3%	82.0%	69.8%	55.5%
	100% Cancerous	27.7%	55.4%	42.4%	100.0%	55.4%	0%	0%	0%	100.0%	0%	17.8%	35.6%	27.3%	100.0%	35.6%
49	50% Cancerous 50% benign	31.8%	53.2%	48.0%	18.2%	18.2%	49.3%	98.6%	99.3%	100.0%	0.3%	38.0%	69.4%	66.3%	47.4%	57.0%
	67% Cancerous 33% benign	33.3%	54.6%	56.3%	21.5%	85.5%	48.8%	97.7%	98.8%	100.0%	0.8%	38.8%	70.0%	71.4%	49.5%	55.2%
	100% Cancerous	28.3%	56.8%	43.4%	100.0%	56.8%	0.0%	0.0%	0%	100.0%	0%	18.2%	36.5%	27.9%	100.0%	36.5%
50	50% Cancerous 50% benign	27.1%	49.1%	40.0%	7.7%	71.0%	46.3%	92.7%	96.1%	100.0%	1.8%	33.9%	64.6%	60.0%	40.6%	46.3%
	67% Cancerous 33% benign	27.7%	49.0%	41.1%	9.8%	65.1%	45.8%	91.7%	95.5%	100.0%	0.9%	34.2%	64.3%	60.5%	42.0%	42.1%
	100% Cancerous	26.7%	53.4%	39.9%	25.3%	53.4%	0%	0%	0%	100.0%	0%	17.1%	34.3%	25.6%	51.9%	34.3

Table 7.7 Comparison of U-Net segmentation results on 20 Whole Slide Images (WSIs) using three distinct models trained on diverse data distributions

Trained Data	Overall Rresults				
	Mean IoU	Accuracy	F1 Score	Recall	Precision
50% Cancerous 50% benign	40.7%	70.4%	68.7%	60.9%	72.3%
67% Cancerous 33% benign	40.7%	70.3%	64.2%	55.6%	53.3%
100% Cancerous	18.8%	37.7%	29.2%	87.5%	37.7%

Table 7.8 Over all results of the models trained on 20 WSI

As shown in Table 7.8 the models that trained on both benign and malignant data has similar overall results. Both models recorded better results than the model trained on only malignant data.

Table 7.9 shows a great improvement in U-Net performance when trained on 45 WSIs. Also, this time U-Net overfitted on benign data when trained on equal distribution of benign and malignant data resulting in horrible performance on malignant data.

WSI Number	Trained Data	Cancerous Area					Benign Area					Overall				
		Mean IoU	Accuracy	F1 Score	Recall	Precision	Mean IoU	Accuracy	F1 Score	Recall	Precision	Mean IoU	Accuracy	F1 Score	Recall	Precision
46	50% Cancerous 50% benign	15.1%	30.3%	14.7%	0%	100.0%	50%	100%	100%	100%	100%	27.5%	55.1%	45.1%	35.7%	100.0%
	67% Cancerous 33% benign	72.2%	85.4%	85.9%	85.1%	93.0%	44.2%	88.4%	93.6%	100.0%	1.4%	62.2%	86.4%	88.6%	90.4%	60.3%
	100% Cancerous	34.8%	69.6%	57.4%	100.0%	69.6%	0%	0%	0%	100%	0%	22.3%	44.7%	36.9%	100.0%	44.7%
47	50% Cancerous 50% benign	11.9%	23.8%	10.3%	0.0%	100.0%	50%	100%	100%	100%	100%	25.5%	51.0%	42.3%	35.7%	100.0%
	67% Cancerous 33% benign	25.1%	41.1%	42.0%	23.9%	93.9%	48.4%	96.9%	98.4%	100.0%	0%	38.1%	72.1%	73.4%	66.3%	41.6%
	100% Cancerous	38.0%	76.1%	66.3%	100.0%	76.1%	0%	0%	0%	100%	0%	16.8%	33.7%	29.3%	100.0%	33.7%
48	50% Cancerous 50% benign	3.1%	6.2%	1.3%	0%	100%	50%	100%	100%	100%	100%	19.8%	39.6%	36.5%	35.7%	100.0%
	67% Cancerous 33% benign	61.8%	93.9%	95.0%	94.3%	97.3%	48.6%	97.3%	98.6%	100.0%	0.2%	52.1%	96.3%	97.6%	98.4%	26.3%
	100% Cancerous	46.8%	93.7%	90.8%	100.0%	93.7%	0%	0%	0%	100%	0%	12.5%	25.2%	24.4%	100.0%	25.2%
49	50% Cancerous 50% benign	9.6%	19.3%	7.7%	0%	100%	50%	100%	100%	100%	100%	24.0%	48.1%	40.6%	35.7%	100.0%
	67% Cancerous 33% benign	43.6%	66.1%	71.4%	59.4%	94.4%	44.8%	89.6%	94.4%	100.0%	0%	44.4%	81.5%	86.4%	86.0%	32.5%
	100% Cancerous	40.3%	80.6%	72.5%	100.0%	80.6%	0%	0%	0%	100.0%	0%	13.8%	27.7%	24.9%	100.0%	27.7%
50	50% Cancerous 50% benign	9.7%	19.5%	7.4%	0%	100.0%	50%	100%	100%	100%	100%	24.0%	48%	40.4%	35.7%	100.0%
	67% Cancerous 33% benign	31.0%	49.9%	55.8%	39.0%	93.9%	45.2%	90.4%	94.8%	100.0%	0.1%	42.3%	82.1%	86.8%	87.5%	19.2%
	100% Cancerous	40.2%	80.4%	72.1%	100%	80.4%	0%	0%	0%	100%	0%	13.8%	27.6%	24.8%	100%	27.6%

Table 7.9 Comparison of U-Net segmentation results on 45 Whole Slide Images (WSIs) using three distinct models trained on diverse data distributions

Trained Data	Overall Results				
	Mean IoU	Accuracy	F1 Score	Recall	Precision
50% Cancerous 50% benign	24.2%	48.5%	41.0%	35.5%	100.0%
67% Cancerous 33% benign	46.9%	82.8%	85.7%	84.5%	41.1%
100% Cancerous	15.7%	31.5%	27.9%	100.0%	31.5%

Table 7.10 Over all results of the models trained on 45 WSI

The over all performance presented in Table 7.10 shows that the best performing model is when trained in more malignant data than benign data.

7.3 Examples

Table 7.11 shows example results of the best performing U-Net models. Training U-Net on more data resulted in more precise and accurate segmentation on the malignant data. But slightly worse results on benign data.

The patch-based classification model results are visually represented in Figure 7.2 and 7.3 . This images depicts the segmentation of the entire cancerous region from a whole Whole Slide Imaging (WSI) by dividing it into patches. The dimensions of Figure 7.2 are 32106x20427 pixel and the dimensions of Figure 7.3 are 30428x25476 pixel To perform the patch-based classification, the images were initially partitioned into patches, and subsequently, a mask was generated by feeding these patches into the model in a sequential manner. Consequently, a mosaic image was formed, as evident in Figure 7.2 and Figure 7.3.

The colors used in this image hold the following interpretations:

- Green: Represents True Positive fields. These correspond to areas correctly identified as cancerous by the model.
- Red: Represents False Positive fields. These areas were mistakenly classified as cancerous by the model when, in fact, they are tumor-free.
- Blue: Represents True Negative fields. These areas were correctly identified as tumor-free by the model.
- Purple: Represents False Negative fields. These regions were erroneously classified as tumor-free by the model, whereas they actually contain cancerous tissue.

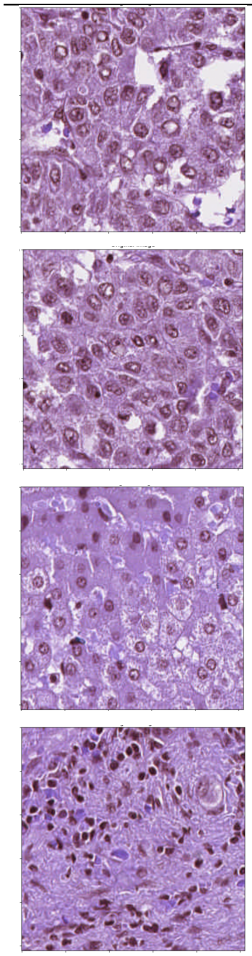
Original Image	Ground Truth Mask	Predicted Mask (Model 1*)	Predicted Mask (Model 2*)
			

Table 7.11 Example results of the trained models
Model 1*: Best performing U-Net model trained on 20 WSI
Model 2*: Best performing U-Net model trained on 45 WSI

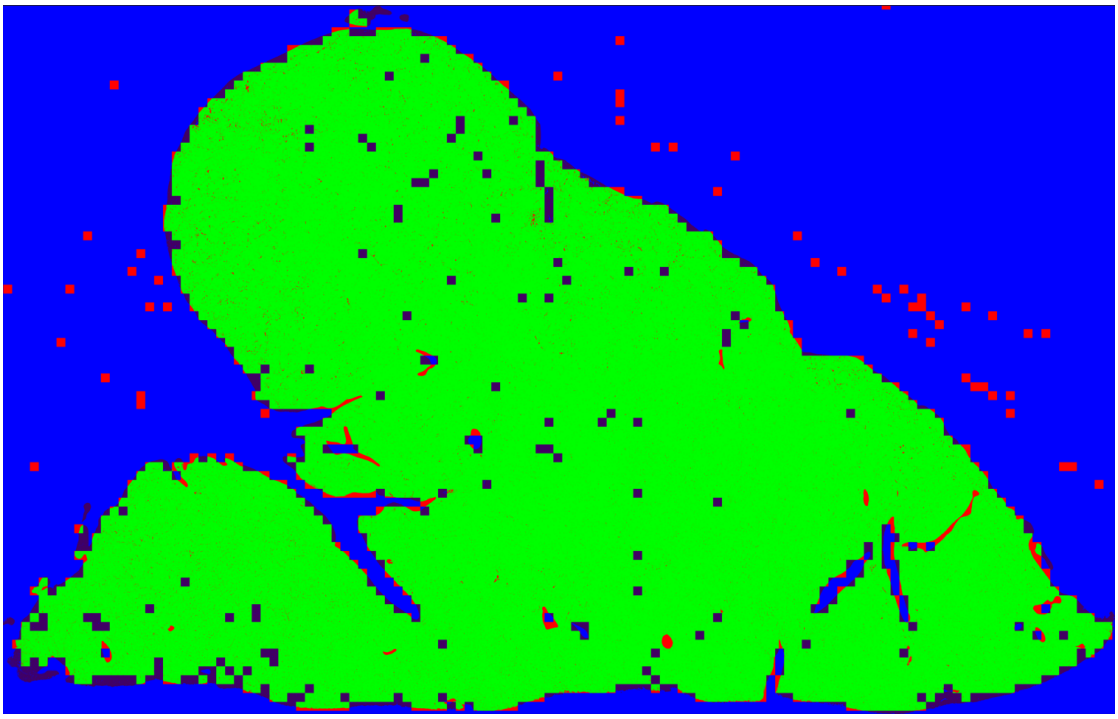


Figure 7.2 Segmentation of Training_phase_2_048

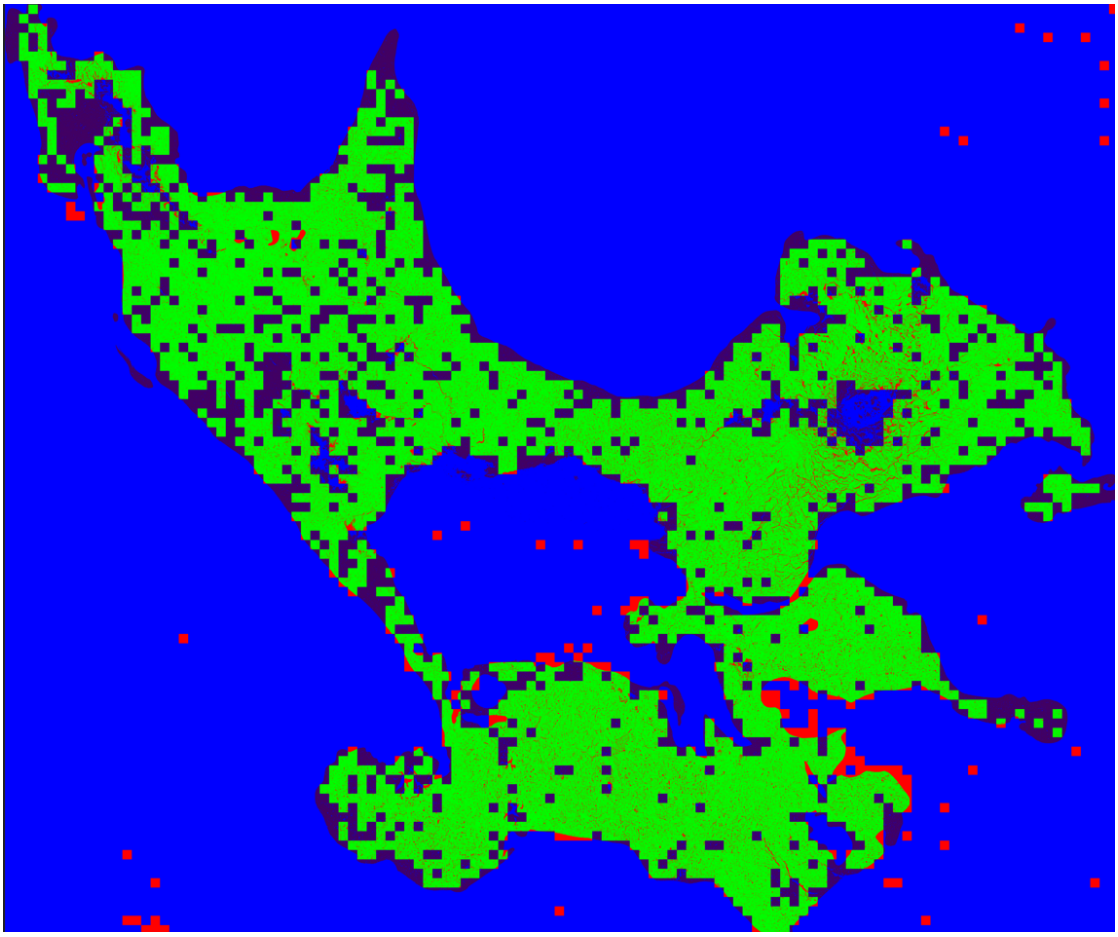


Figure 7.3 Segmentation of Training_phase_2_050

8
Performance Analysis

The results were tested using 3 different models: CNN and ResNet for patch-wise classification and U-Net for pixel-wise classification (segmentation). The CNN and U-Net models were trained on 2 different amount of data: 20 WSIs and 45 WSIs. In this case, it is an expected result that the success will increase with the increase in the number of data. However, the increase in the amount of data made the test and train processes more recourse consuming in terms of time and hardware. The other method, ResNet18, was implemented by fine-tuning the trained version of the ImageNet dataset that contain millions of images. For the patch wise-classification ResNet18 achieved better performance than CNN. Especially in the CNN model trained in Training_phase_2_047, accuracy score increased close to 20% and f1 score increased up to 3 times.

It has been observed that training the models on WSIs with a balanced distribution give better performance. The reason for this is that the models trained on unbalanced distributions have a chance to overfit, resulting in bad performance.

Regarding U-Net model, different distributions and data size were tested to find the best performing model. As shown in Table 7.7 and Table 7.9 when training U-net on only cancerous data the model becomes overfitted towards cancerous data resulting in 0 accuracy on benign data.

Also, Table 7.9 shows that when training U-Net on 45 WSIs with 50% benign and 50% malignant data the model becomes biased towards benign data. This can be because of the huge amount of benign patches in the 45 WSIs. It is crucial for U-net to balance the data used for training in order to get satisfying results.

The best performing U-Net model is the one trained on 45 WSIs using 67% cancerous and 33% benign data resulting in 46% Mean IoU, 82.8% Accuracy, and 85.7% F1 score.

9
Conclusion

Cancer segmentation and classification is very hard and complex task since it needs a lot of data and a very powerful hardware. In this project we tried to achieve this task by training 3 different models on 45 WSIs.

Considering the results obtained in the patch-wise classification method, increasing the data has improved the performance of the models. Also, higher performances were obtained when the regions labeled as partially cancerous were evaluated as cancerous. This shows that the model has learned the cancerous areas well and that the patch can be considered as cancerous even if it is less than 80% cancerous. In order to better solve this problem, a more optimal value can be found by trying different values as threshold other than 80%.

It is observed that fine-tuned ResNet architecture gives the best result. But, the disadvantage of ResNet is that it takes so much time to be trained and tested. In order to increase the success of this model, the last CNN block of the model can be unfreeze and trained on the new data.

Another solution is to train only the classification layer using a fine-tuned version of ResNet18 over medical images. The model were tested on 5 WSIs. As can be seen from the results' tables **??**, 7.5 and 7.6, while some images achieve very high success, some images have low success. This makes it difficult to measure the overall success of the model. For a better evaluation, k-fold cross validation method can be used.

Moreover, it is feasible to enhance Figure 1 and Figure 2, which visualize the segmentation results, by applying post-processing techniques utilizing image processing methods. Specifically, improvements can be made by addressing the spaces between the segmented regions. One approach is to employ low-pass filters to eliminate noise present in the images. These filters can effectively smooth out the images, reducing any unwanted artifacts or irregularities that may be present. By implementing such post-processing techniques, the overall quality and clarity of

the visualized segmentation results can be enhanced, leading to a more accurate representation of the underlying information.

Regarding Pixel-Wise classification, U-Net achieved a satisfying performance when trained on a balanced distribution and big around of data. In order to achieve better performance it is possible to increase the amount of data by including another data set into the training or applying data augmentation on the current dataset.

References

[1] C. Gaggero, *Promoting cancer early diagnosis*. [Online]. Available: https://www.who.int/activities/promoting-cancer-early-diagnosis.

[2] M. Cui and D. Y. Zhang, "Artificial intelligence and computational pathology," *Laboratory Investigation*, vol. 101, no. 4, pp. 412–422, 2021.

[3] V. Baxi, R. Edwards, M. Montalto, and S. Saha, "Digital pathology and artificial intelligence in translational medicine and clinical practice," *Modern Pathology*, vol. 35, no. 1, pp. 23–32, 2022.

[4] A. Patel, "Benign vs Malignant Tumors," *JAMA Oncology*, vol. 6, no. 9, pp. 1488–1488, Sep. 2020, ISSN: 2374-2437. DOI: 10.1001/jamaoncol.2020.2592. eprint: https://jamanetwork.com/journals/jamaoncology/articlepdf/2768634/jamaoncology_patel_2020_pg_200003_1599155176.12055.pdf. [Online]. Available: https://doi.org/10.1001/jamaoncol.2020.2592.

[5] J. C. Choi, P. Park, S. Y. C. Chun, W.-K. J. Jeong, and K. Lee, *Paip 2019 - grand challenge*, Apr. 2019. [Online]. Available: https://paip2019.grand-challenge.org/Dataset/.

[6] B. V. Janssen *et al.*, "Artificial intelligence-based segmentation of residual tumor in histopathology of pancreatic cancer after neoadjuvant treatment," *Cancers*, vol. 13, no. 20, p. 5089, 2021. DOI: 10.3390/cancers13205089.

[7] B. V. Janssen *et al.*, "Artificial intelligence-based segmentation of residual tumor in histopathology of pancreatic cancer after neoadjuvant treatment," *Cancers*, vol. 13, no. 20, p. 5089, 2021. DOI: 10.3390/cancers13205089.

[8] P. Bilic *et al.*, "The liver tumor segmentation benchmark (lits)," *Medical Image Analysis*, vol. 84, p. 102 680, 2023, ISSN: 1361-8415. DOI: https://doi.org/10.1016/j.media.2022.102680. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841522003085.

[9] F. Ettlinger, F. G. Gruen, and S. Schlecht, *Competition*, P. C. Christ, Ed., Jun. 2017. [Online]. Available: https://competitions.codalab.org/competitions/17094#learn_the_details.

[10] H. Rahman, T. F. Bukht, A. Imran, J. Tariq, S. Tu, and A. Alzahrani, "A deep learning approach for liver and tumor segmentation in ct images using resunet," *Bioengineering*, vol. 9, no. 8, p. 368, 2022. DOI: 10.3390/bioengineering9080368.

[11] E. Vorontsov, A. Tang, C. Pal, and S. Kadoury, "Liver lesion segmentation informed by joint liver segmentation," *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 2018. DOI: 10.1109/isbi.2018.8363817.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, Springer, 2015, pp. 234–241.

[13] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: Analysis, applications, and prospects," *IEEE transactions on neural networks and learning systems*, 2021.

Curriculum Vitae

FIRST MEMBER

Name-Surname: Emre ARSLANOĞLU
Birthdate and Place of Birth: 06.09.1999, Samsun
E-mail: emre.arslanoglu@std.yildiz.edu.tr
Phone: 0501 069 19 65
Practical Training: Doğuş Teknoloji Şirketi İleri Analitik Departmanı, Anadolu İsuzu Şirketi Araştırma Geliştirme Departmanı

SECOND MEMBER

Name-Surname: Muhanad TUAMEH
Birthdate and Place of Birth: 22.06.2021, Şam
E-mail: Muhanad.tumah@std.yildiz.edu.tr
Phone: 05385238952
Practical Training: INOSENS Yapay Zeka Departmanı, Wolves İnteractive Yazılım Departmanı

Project System Informations

System and Software: Windows İşletim Sistemi, Python
Required RAM: 16 GB
Required Disk: 512 GB