AI-enabled surgery: Enabling da Vinci Xi Robot with Live Cancerous Tumour Detection

Rohan Ibn Azad

Master of Research Y2

School of Engineering Macquarie University

October 25, 2021

Supervisor: Dr Mohsen Asadnia

Co-Supervisor: Prof. Subhas Mukhopadhyay

STATEMENT OF CANDIDATE

I, Rohan Ibn Azad, declare that this report, submitted as part of the requirement for the award of Master of Research in the School of Engineering, Macquarie University, is entirely my own work unless otherwise referenced or acknowledged. This document has not been submitted for qualification or assessment at any academic institution.

Student's Name: Rohan Ibn Azad

Student's Signature:

Date: 25th October 2021

Table of Contents

Abstract

Deep learning has proved successful in Computer Aided Detection in interpreting ultrasound images, CT scans, identifying COVID infections, identifying tumors from ultrasound, Computed Tomography (CT) scans for humans and for animals. Currently, only experienced surgeons can identify tumors in patients with kidney cancer using ultrasound and Indocyanine Green (ICG) with Fluorescence Imaging which may come with error. Therefore, this project proposes applications of deep learning in detecting cancerous tissue inside patients via laparoscopic camera on da Vinci Xi surgical robots. The proposed algorithm can help the surgeons to detect cancerous tumors from fatty tissue and non-cancerous tissue with 84% accuracy during the surgery which is extremely beneficial to ensure all the cancerous tumors are removed. The process is carried out via object detection techniques which draws bounding boxes and shows the probability for that region to be cancerous tissue, non cancerous tissue or fatty tissue, which is the primary goal of the project. The project compares between optimized AlexNet, VGG-16, YOLOv3, YOLOv4 to work out the best algorithm with tuned hyperparameter to detect cancerous tissue during surgery. Analysing images, the final mAP for object detection was 0.974 and for classification, the accuracy was 0.84.

1. Introduction

1.1 Information on prostate cancer and kidney cancer

"Cancer" is a word that groups a large number of diseases that is caused by rapid, uncontrolled cell division. If not detected in early stage, cancer can spread to other organs and tissue [1]. With Mitosis process, depending on the type of the cell, normal cells have a specific life cycle. DNA inside cells have instructions that control the life cycle of cells and the old cells are replaced by new cells. Even though DNA instructions often get mutated, they can get corrected by cells. The cells not being able to correct the mutations results in cancerous cells as show in Figure 1 [1].

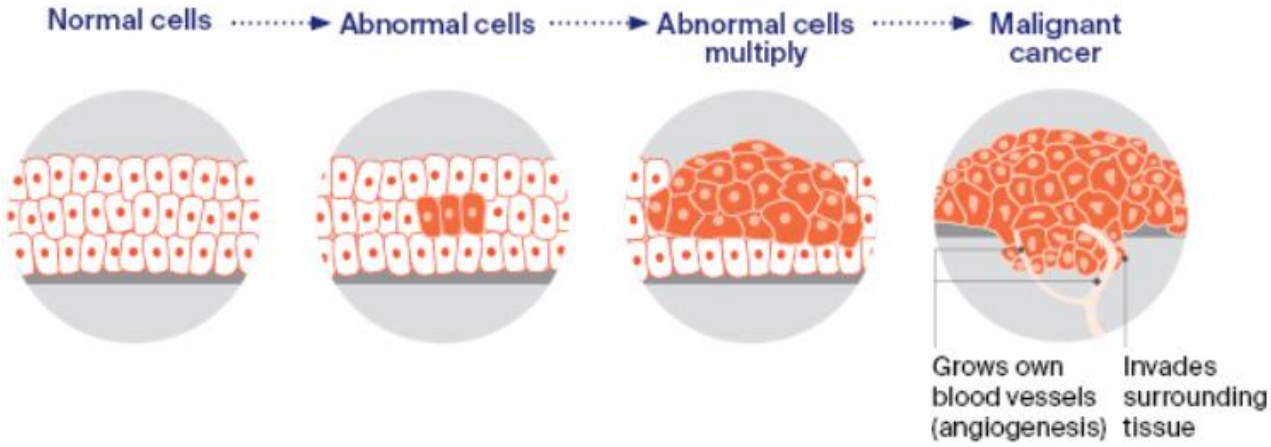


Figure 1: Normal cells becoming mutated and cancerous [2]

One of the deadliest types of cancer in men is prostate cancer [2]. In the male anatomy, a small gland that surrounds the urethra below the bladder is called prostate. Prostate cancer is when uncontrollable cells division starts occurring in the prostate gland [3]. Prostate cancer can develop either on the gland cells or other cells inside the prostate [4] as shown in Figure 2(a). The gland cells are called –

1. Basal cells.
2. Luminal cells.

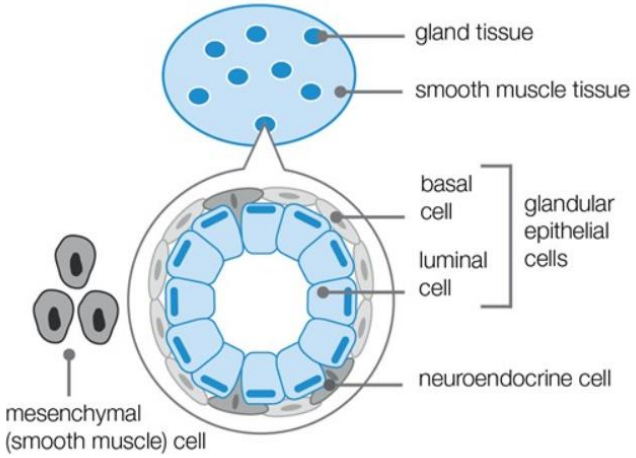


Figure 2 (a): Tissue structure of prostate [5]

Despite all the progresses in this domain, there is still no permanent treatment for cancer. Among the available semi-temporary solutions such as- radiation therapy, hormone therapy, palliative care, surgery, the surgery is the most common option for most men [5] [6].

The surgical procedure is called radical prostatectomy. The surgery aims to remove all cancerous tissue which has some side effects such as the patient may become impotent [6].
Like prostate cancer, male population has the greater possibility of getting affected by kidney cancer, more than 2-3 times of female [7] [8]. Since kidney cancer grows on the outer side of kidney, it has a greater chance of spreading to other organs such as lung as shown in Figure 2(b).

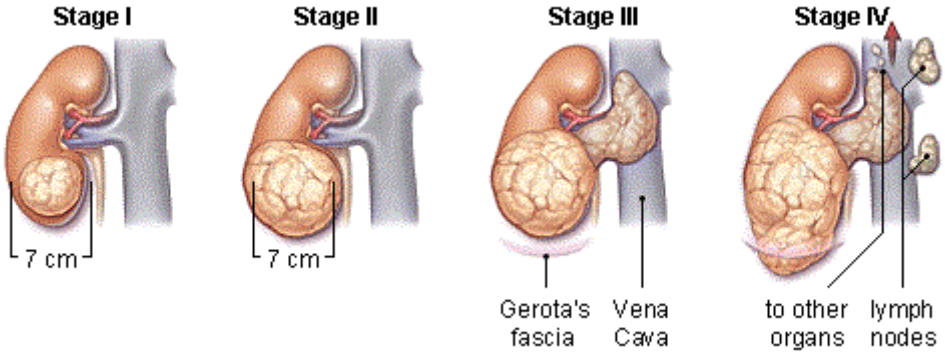


Figure 2(b): Kidney cancer stages [7]

Unlike prostate cancer, kidney cancer can be treated by cutting out the malignant tumors on kidney. The process is called partial nephrectomy.

1.2 Laparoscopy

Prostate cancer is detected using diagnosing tools such as ultrasound, Computed Tomography (CT) scan, Magnetic Resonance Imaging (MRI) scan. When these non-invasive methods fail to give reliable diagnosis, the other option is to use laparoscopy [9].

But owing to laparoscopy's unstable camera platform, limited mobility of the camera, images in two-dimensional representation, uncomfortable operating arrangement for the surgeon raised the need for robotic surgery. The da Vinci Xi has got safety features in place. When the robotic arms are connected to the First entry Instrument, the arm that carries the camera locks in place. The surgeon can adjust their sitting position to be more ergonomic in front of the console.

1.3 da Vinci Xi Surgical Robot

da Vinci Xi is a minimally invasive robotic assisted surgical system. The system is developed, manufactured, marketed by the American company Intuitive Surgical Inc [10]. The term "robotic" can be misleading. The robot does not perform the surgery rather the surgeon guides the robot via a console.

(a) (b) (c)

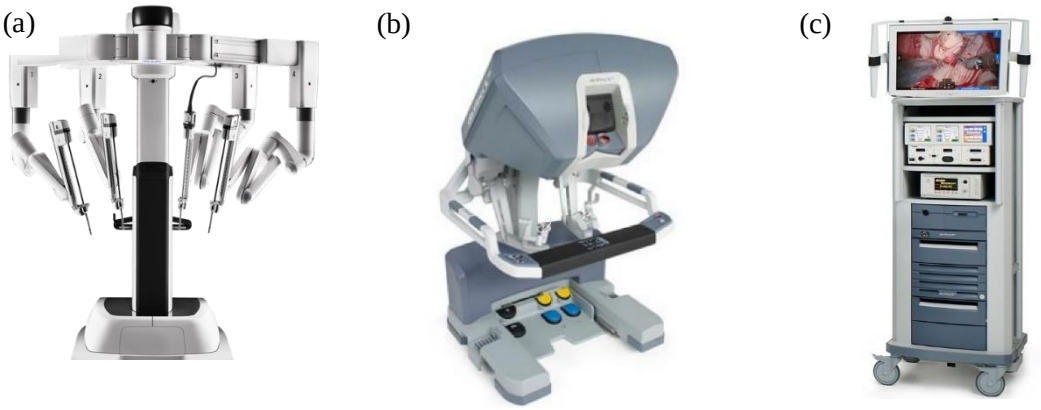


Figure 3: Components of the da Vinci Xi robot (a) Patient cart (b) Surgeon console (c) Vision cart [10]

In Figure 3, as shown, the da Vinci Xi Surgical Robot has four arms, arm 1 in patient cart carries a laparoscopic camera, the other 3 arms carry various cutting tools for performing surgery. The surgeon controls the robotic arms via the console and gets a 3D view of the operation. The Vision Cart works as a second screen with 2D view for the nurses and assistant surgeon in the operation theatre and also works as the central processing unit for the system. [10]

da Vinci Xi gives surgeons the opportunity to perform Minimally Invasive Surgery (MIS) [10]. With MIS, a surgeon makes small incisions each of about 1-2 cm and inserts the patient cart (Figure 3 (a)) arms rather than making one large cut to the region of interest of the body [11] [12]. The arms carry small cutting tool, gripping tools, laparoscopic camera as shown in Figure 4 (a) [13]. In Figure 4(b), the laparoscopic camera that works as the surgeon's eye during surgery is shown. The output from the laparoscopic camera is shown in 3D form in surgeon's console and in 2D form in the vison cart. The first entry instrument as shown in Figure 4(c) is the part that works as a connector between the incision on the patient and the robotic arms. The first entry instrument first goes into the patient, then the robotic arm is attached to the instrument and then the camera, cutting tools go through the first entry instrument into the patient. Once the arm specifically designed for carrying the camera gets in contact with the instrument, the arm locks in place and cannot be moved during surgery. This works as a safety feature for the robot.

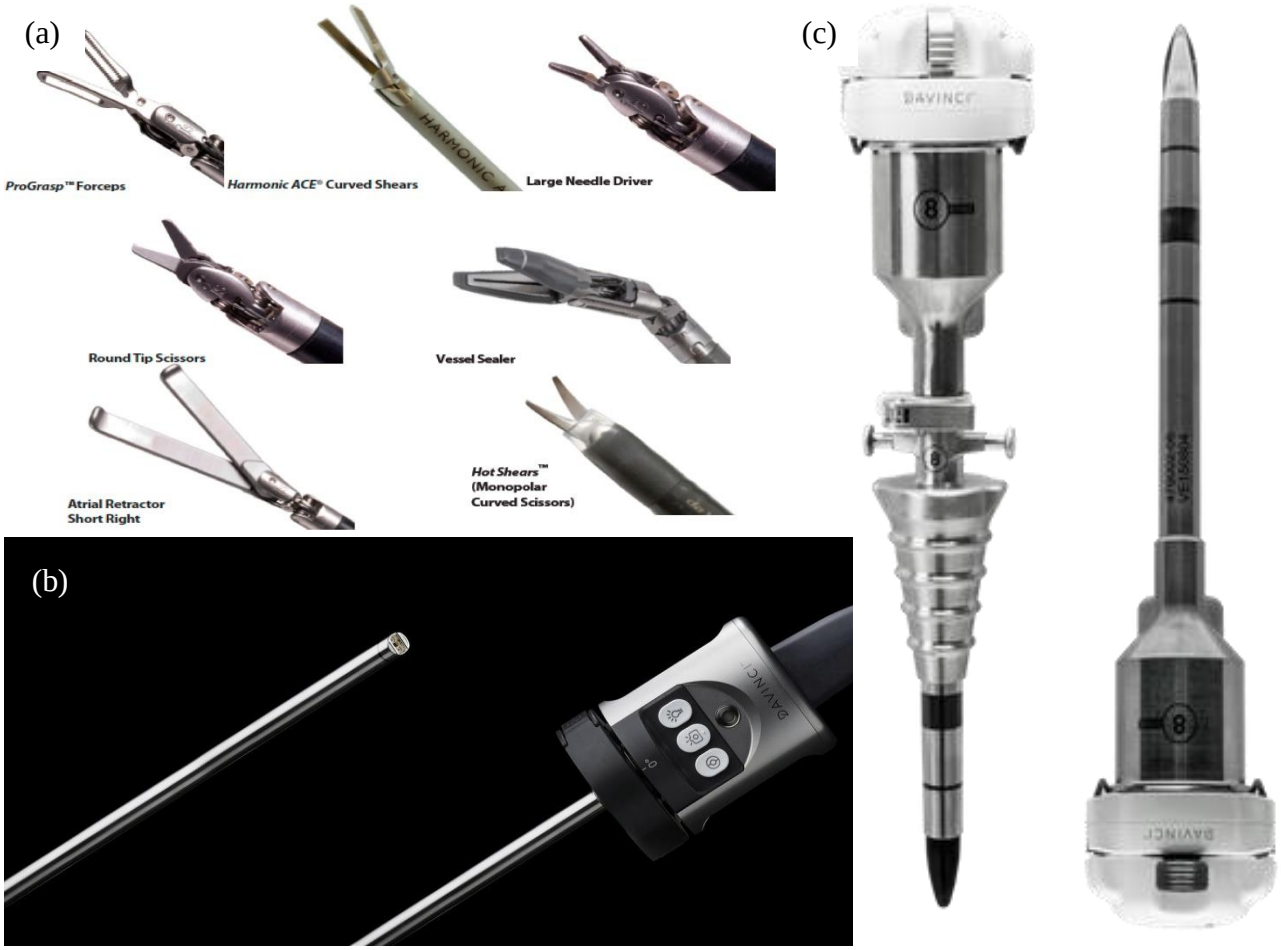


Figure 4: (a) Cutting and gripping tools [13], (b) Laparoscopic camera [14] (c) First entry Instrument [15]

1.4 da Vinci Xi image capture

The da Vinci Xi Vision (Figure 4(b)) can capture images in 2 ways. One is simple laparoscopy and the other one is using Ultrasound (US) probe to create a 3D reconstruction of the region of interest on an anatomical visualization service called Iris.

During surgery it is challenging for the doctor to distinguish between cancerous and non-cancerous cells [6]. From [9], one of the diagnosing techniques for prostate cancer is MRI. But MRI is not accessible intra-operatively. da Vinci Xi uses Trans Rectal Ultrasound (TRUS) probe to create a 3D representation of the region of interest as shown in Figure 5 (a).

(a) (b)

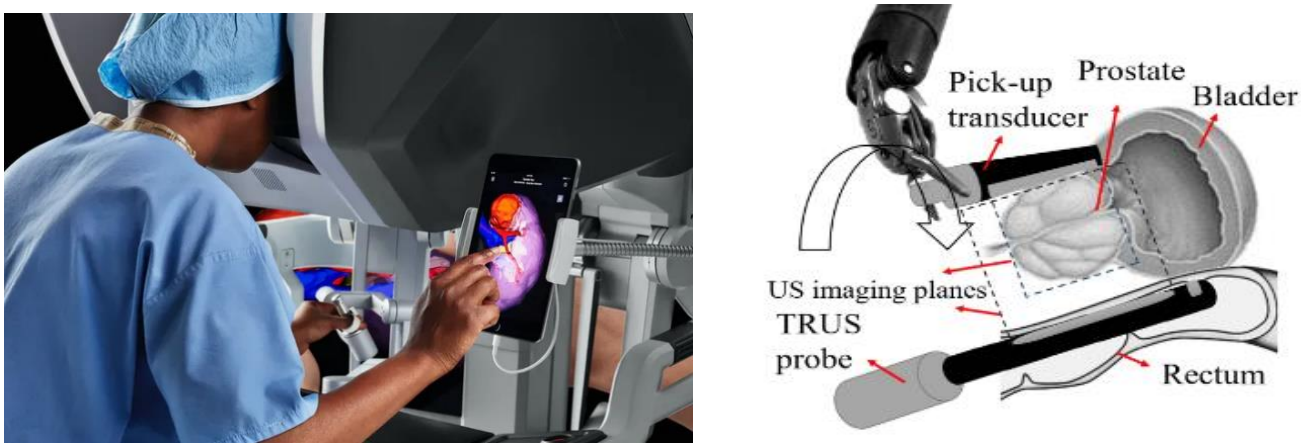


Figure 5: (a) Surgeon looking into the console and making markings on the 3D reconstruction on Iris [14], (b): Placement of TRUS pick-up probe for 3D reconstruction [6]

In Figure 5 (b), the TRUS works as a light source for the photoacoustic effect. The non-ionizing laser released from the TRUS is absorbed by the tissue and converted into heat. Different tissues react to it differently as they have different thermal expansion coefficient and in turn releases different ultrasonic emission that is captured by the pick-up probe. The signal picked up by the pick-up probe is reconstructed into 3D shape by using Delay- and-Sum (DAS) and iterative Deconvolution-based PA Reconstruction with Sparsity Regularization (iDPARS).

1.5 Lack of Artificial Intelligence in the camera functionality

The image capturing technique mentioned in section 1.4 is not full proof as the surgeon will have to look into the console, mark the regions in the Iris tablet, look back into the console. The process is not very intra-operative. Also, kidney cancer is very difficult to detect at an early stage [16]. If there was a technology that would show the surgeon the cancerous regions in bounding boxes in the console, that would save the surgeon a lot of time and hassle of switching from console to Iris tablet and vice versa.

da Vinci Xi enables robotic surgery using small incisions which can significantly help the surgeons with cancerous tumor removal surgery. da Vinci surgical robot was first commercialized by intuitive in 2000. In 2014, intuitive released da Vinci Xi [59]. The da Vinci Xi robot has 3 parts, the patient cart, the surgeon console and the vision cart. The patient cart has 4 arms to be used during the surgery, the surgeon controls the robotic arms through the console, the vision cart works as the CPU for the system and works as the second screen [60]. Currently the surgeons rely on their experience to identify the tumors. Once the tumor's location has been approximated, da Vinci Xi provides intra operative ultrasound and Indocyanine Green (ICG) with Fluorescence Imaging to further assist the surgeon. Intra operative ultrasound shows the depth of the tumor and makes a 3D

reconstruction of the organ on a tablet beside the surgeon's console. Injecting ICG and turning on fluorescent light makes the kidney green and the tumor grey. But if the tumor location cannot be identified then intra operative ultrasound will not work. Not injecting ICG in the correct dose will either make the whole field of view green or will not change color. ICG also comes with side effects which makes it necessary to keep the injection of ICG minimum [61]. There had been remarkable progress made in medical applications of image processing due to the availability of open source large scale annotated datasets. The applications include both pre and post- operative diagnosis. In 2015, Support Vector Machine (SVM) was the most reliable classifier. Papers presented before Chung et al [62] only considered one slide from each MRI scan. Chung for the first time considered the spatial information contained in 3D voxels in the MRI scans. After 2015, when Deep Neural Networks gained some insights as to how they work owing to the work of Zeiler et al [63], Convolutional Neural Networks became popular for image classification. Shin et al [64] made use of publicly available CT images for thoraco-abdominal lymph node detection and interstitial lung disease classification. Pantanowitz [65] et al fulfilled the need for computer assisted diagnostics of prostate core needle biopsies (CNBs) by developing an algorithm that takes input as hematoxylin and eosin (H&E) stained slides outputs the result with 0.997 AUC. Deep learning was also used for detecting cancer in animals [66], agricultural greenhouse detection [67], analyzing traffic load distribution on a bridge [68], airplane detection [69], hand gesture recognition [70], automatic vehicle inspection [71], license plate recognition [72]. All these works presented here, only focused on pre and post-operative diagnosis using Magnetic Resonance Imaging, Computer Tomography scans, Ultrasound Images. None of the papers consider real time surgical images to identify tumors. This project will address this issue and propose using Convolutional Neural Network (YOLOv4 at first, then optimized VGG-16) for giving the surgeons a second opinion during real time tumor removal surgery.

This project suggests incorporating a trained computer vison algorithm to recognise cancerous tissue and draw bounding boxes around those. The questions that the thesis addresses are – can the developed optimized algorithm identify cancerous tumors real time during minimally invasive surgery with high level of accuracy? Is it capable of identifying tumors on any organ? Is the algorithm able to provide reliable results on any related dataset?

2. Literature review

2.1 Introduction

This section covers past work done in Convolutional Neural Network (CNN) including the development of some algorithms and evolution of the application pattern of those algorithms. The

section covers how the development of different CNN helped in improving the performance of cancer detection.

A paper on this project was published in International Journal on Smart Sensing and Intelligent Systems in volume 14, issue 1, pages 1-16. The published paper summarises everything presented in this thesis in a more compact form.

2.2 Development of Convolutional Neural Network

The concept of Convolutional Neural Network came from the concept of the visual cortex of the brain [17]. According to D. Hubel and T. Wiesel, [18] [19] each of the neurons in the brain's visual cortex have their local receptive field. The receptive field of one neuron overlaps with the neuron next to it and by overlapping with each other, the entire visual field is created.

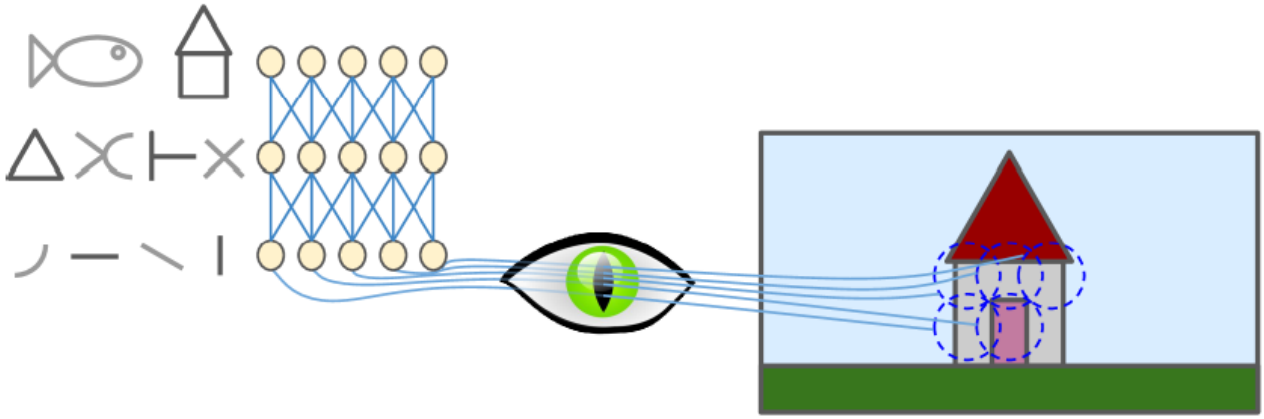


Figure 6: Simple illustration of the visual cortex [17]

In Figure 6, the blue dashed circles on the image of the house are the receptive fields of the neurons. The blue dashed circles overlap with each other. The yellow circles represent each neuron in each layer. The Figures next to each of the layers show that the first few layers interpret the building blocks of the image such as lines, edges and eventually builds the entire image at the final layer.

The study of the visual cortex inspired K. Fukushima to create a self-organizing neural network for pattern recognition in 1980 [20]. In 1998, Yann LeCun [21] first came up with the idea of convolutional layers, pooling layers and proposed LeNet-5 handwritten digit recognizer architecture.

There was not a lot of research done on computer vision since 1998 until 2012, because the functionality of the hidden networks in deep neural networks was still a black box and significant investment went into improving Support Vector Machines (SVMs) in machine learning. In 2012, A. Krizhevsky [22], came up with AlexNet where the network architecture was larger and deeper than LeNet. ImageNet [23] 2012 challenge was won by using AlexNet. In 2013, deconvolutional neural network [24] was worked out to Figure out the inner working of the hidden layers in CNN to further

optimize the network architectures. Matthew D. Zeiler proposed a way to visualize the filters of each hidden layer to tune the hyperparameters accordingly and optimize the network architecture.

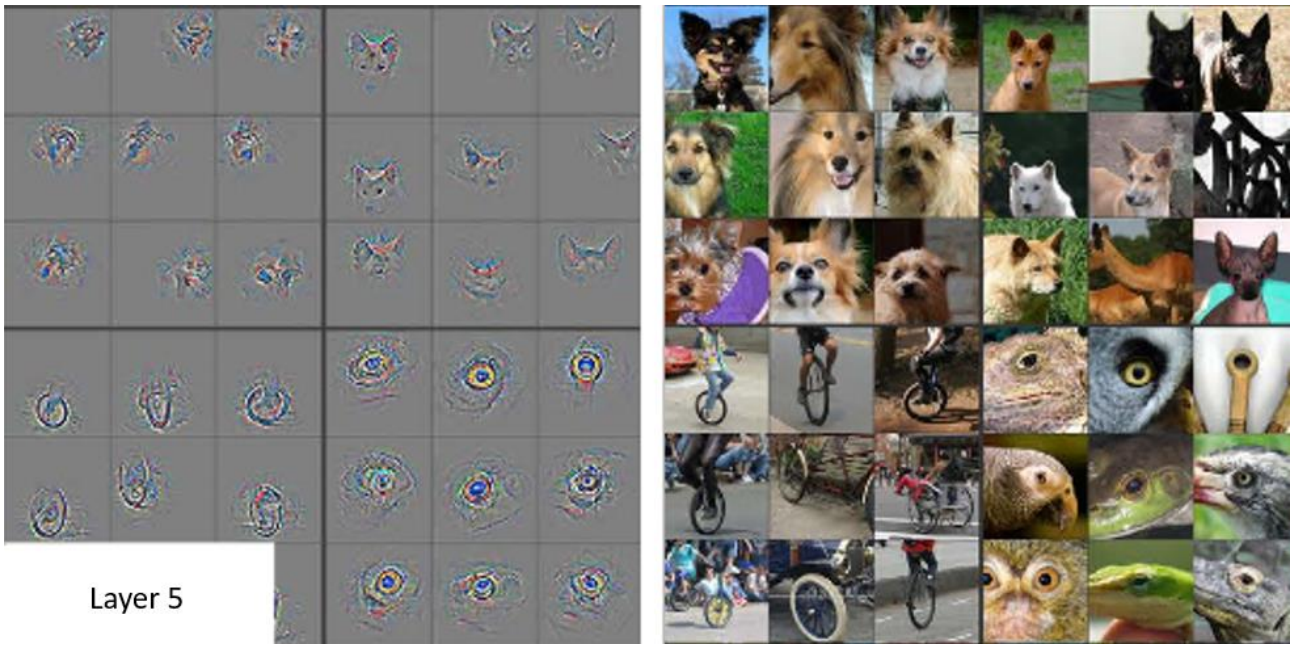


Figure 7: Filters of Layer 5 in AlexNet after training the optimized network [24]

From Figure 7, the visualization of the filters of the hidden layer ensures that the network architecture had been optimized for classification. Here for each of the filters in layer 5 (the grey portion in the left of Figure 7), the corresponding image patch (the coloured portion in the right of Figure 7) is also shown. The same technique can be employed to tune the hyperparameters of any classification network architecture.

From [24], it was found that having a deeper network architecture improved the performance of the classifier. Increasing the layers in AlexNet, in 2014, VGG-16 network architecture was proposed by K. Simonyan and A. Zisserman [25], which further improved the performance of AlexNet on the ImageNet dataset. As the availability of training data is increasing and the computational power is increasing, more and more optimisation of image classification and object detection is being done.

2.2 Pre-Fully Convolutional Network Application in surgery

One of the key applications of deep neural network is biomedical image processing. Doctors, radiologists now prefer to have a second opinion on diagnosis coming from the output of CNNs. In [26], the authors discussed about how AI helped a doctor from making a wrong diagnosis of cancer for a young girl. After a young girl who had been treated for medulloblastoma came back for a routine follow up at New York University (NYU), it was found that the ailment had returned, and it was confirmed by biopsy. The doctor who was looking after the case, before signing the chemotherapy forms, used the results from biopsy and inputted them into an Artificial Intelligence (AI) classification

system. The prediction from the AI system came out to be a different kind of cancer (glioblastoma). This allowed the doctor to make a correct radiation treatment plan and prescribe different drugs [26].

Referring to A. Chung [27], L. Pantanowitz [28], Y. Wang [29], research work had been done to detect prostate cancer for diagnosis using Magnetic resonance Imaging (MRI) scans and Ultra-sound scans. On 3D multi-parametric MRI, A. Chung [27] discussed about using radiomics-driven prostate cancer detection that uses spatial information of the 3D voxels rather than focusing on individual voxels as the older radiomics-driven prostate cancer detector used to do. The author extracted radiomics features from multi-parametric MRI using a quantitative radiomics feature model. The author used a Support Vector Machine (SVM) classifier to get initial detection of cancer and combined the output from SVM with radiomics-driven conditional random field (RD-CRF) framework to get the final detection. Even though this method achieved accuracy more than its predecessors, it is very trivial and quite slow.

Y. Wang applying [29] Fully Convolutional Network (FCN) in multi parametric MRI paved the way for application of Fully Convolutional Network in medical field. He tried to apply V-Net, a network architecture specifically designed to capture cancerous cells from 3D MRI and segment those. V-net was developed by F. Millerati [30]. Instead of using densely connected neural networks at the top of convolutional layers like traditional CNN, V-Net uses fully convolutional network. This preserves the spatial information of the images and allows for detection to take place. The author took a two-stage detection approach where in the first stage the prostate was segmented from the background and in the second stage the tumor was detected. Dice coefficient was used as an evaluation metric. The author achieved a dice coefficient of 0.8935.

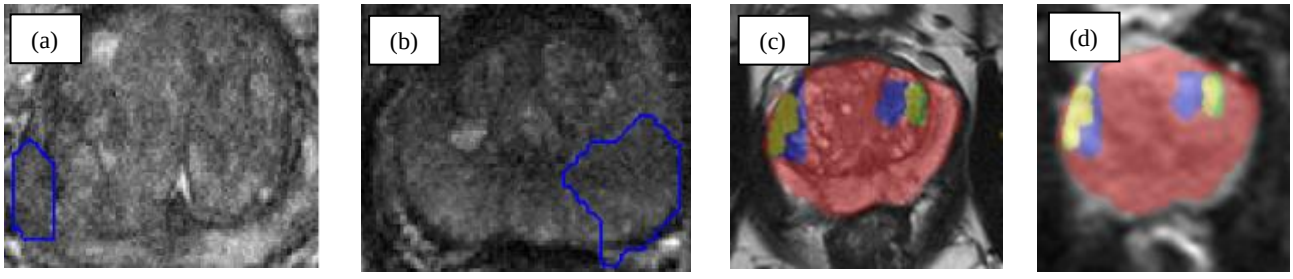


Figure 8: (a), (b) shows the prostate cancer identification by using RD-CRF. (c), (d) shows prostate cancer identification by using FCN [27] [29]

From Figure 8, it is quite easily distinguishable that FCN performed better than RD-CRF.

L. Pantanowitz [28] was the first who besides developing an algorithm for prostate cancer detection and then afterwards assigning a Gleason score, also applied it for clinical use. The algorithm was applied on Core Needle Biopsy (CNB). The author did not specify what algorithm was used in this case. It was only mentioned that the used algorithm's base was a multilayered convolutional neural

network. The author also mentioned a Gradient Boosted classifier as a segmentation approach for separating the tissue from the background was used first and then 3 CNN are ensembled to detect cancer from the tissue areas.

Researchers such as G. Aly[31], Y. Wang, H. Murat [32], N. Tangri [33] had suggested to use computer aided diagnose (CAD) to help doctors and radiologists cope with the large amount of big data that is being generated. Since the main purpose of the project is detecting cancerous tumors, besides detecting kidney tumors, the project also covers past work in breast cancer, skin cancer, canine mammary carcinoma detection, COVID-19 detection, prostate cancer.

T. Grigore had used Inception v3 neural network on 3D rendered CT scans to predict the staging of kidney cancer (the stages mentioned in (Figure 2(b)). The images were cropped by using ImageJ making sure the cropped portion included kidney cancer. An AUC score of 0.90 for the test set was achieved. In canine mammary carcinoma, mitotic count from Whole Slide Images (WSI) of canine breast is analysed to be used in human breast cancer research [34]. Inaccurate mitotic count can lead to wrong diagnosis. The WSIs that are available for human breast cancer do not contain annotations for the entire WSIs. Keeping in mind the need of an algorithm to detect mitotic count in WSIs, several challenges including MITOS 2012 dataset had been released. The F1 score from the model was 0.66 and the result from the model was flawed and the algorithm was not considered state-of-the-art anymore as the algorithm had been trained and tested from the same data [34]. The author suggested using a combination of RetinaNet and then ResNet to increase the efficiency of identifying mitotic counts in 21 WSIs of Canine Mammary Carcinoma. The author's proposed method achieved a F1 score of 0.791, which is a significant improvement from the supposed to be state-of the-art model for identifying mitotic counts of Canine Mammary Carcinoma.

Convolutional Neural Networks helped immensely during the global pandemic of 2020 as well. CNN was used in effectively and quickly identifying COVID-19 cases where the patients were asymptomatic. Brunese [35] discussed about the new deadly Corona Virus that was spread throughout the world and caused a global pandemic in 2020. According to Huff [36], sometimes patients infected with COVID-19 did not show any symptoms and as a result they spread the virus without them being aware of the situation. The method for testing for COVID-19 was swab test but it was only carried out on people who were showing symptoms of COVID-19. In the paper, the author suggested using Convolutional Neural Networks (CNN) on X-ray images in 3 steps-

- Using VGG-16 (Visual Geometry Group) to determine if a chest X-ray belonged to a healthy patient or to a pulmonary disease affected patient.

- Using VGG-16 to determine if the pulmonary disease affected patient was suffering from pneumonia or COVID-19.
- If the patient was suffering from COVID-19, it would be identified using a visual localization technique with transfer learning. This is otherwise known as GRAD-CAM (Gradient based Class Activation Mapping) [37]

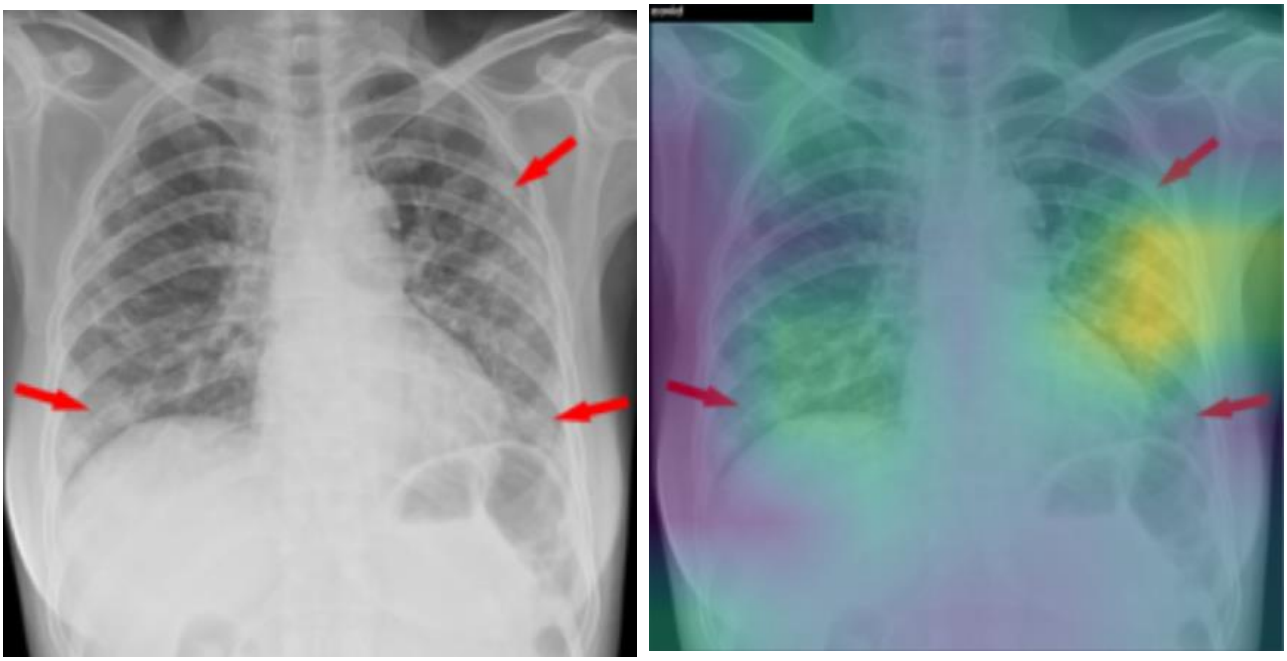


Figure 9: On the left is the original chest X-ray with COVID-19. The red arrows are labelled by a specialist indicating those are the portions with COVID-19 infection. On the right, the ML algorithm with GRAD-CAM showing highlighted region was focused to give the end results as COVID-19 positive [35]

Here, VGG-16 with GRAD-CAM (Figure 9) achieved an overall accuracy of 0.97 and COVID-19 detection took approximately 2.5 seconds. In the medical applications mentioned above, most of those were done in microscopic level using ultrasound and MRI. Also, in all of those medical applications, the object detection diagnosis was done before surgery.

Researchers had used the detection algorithms mentioned above for various purposes including but not limited to breast masses detection in full-field digital mammograms, detecting and distinguishing between pulmonary disease and Corona Covid-19 detection by using X-ray images [31], detecting skin lesion in dermoscopic images [20]. Since the primary objective of this project is related to object detection, research papers outside of medical object detection had also been analysed. Agricultural greenhouse detection [38], analysing traffic load distribution on a bridge [39], airplane detection [40] are some of those.

2.3 Post Fully Convolutional Network application in surgery

The researches discussed in "Pre-Fully Convolutional Network Application in surgery" section used classification to decide if the images had any sort of lesions present in those. If the lesions were

present they localised the lesions in the images. If the process is done for multiple lesions in the image, the process is called object detection. Until 2015, a sliding window object detection approach was used for CNN. A CNN was trained to classify and locate objects and then the CNN window was used to slide across the entire image. The problem with this method is that since different objects can have different sizes, so n x n (here, n is a number belonging to an integer set) sliding window would not always be the suitable size for every object detection. So, the process would have to go through varying sizes of sliding window regions. Also, as the window was being taken across the entire image, the same objects would get detected couple of times. As the sliding window regions would have to be varied here, and as the same objects would get detected multiple times, this process would be slow because the CNN would have to run many times. [17]

Fully Convolutional Network (FCN) is the solution to this. In 2015, Jonathan Long [41] came up with the idea of FCN. In traditional Convolutional Neural Network architecture, the convolutional layers are flattened and then fed into a densely connected neural network. Flattening the convolutional layers eliminates spatial information and it is not possible to do object detection in traditional Convolutional network with densely connected neural network on top. In Fully Connected Networks the dense layer on top of the last convolutional layer in the network is replaced with 1 x 1 convolutional layers. This allows the convolutional layer to retain spatial information and detection action can be performed. Another advantage of using fully convolutional network is that as long as the input image size is bigger than the filter or kernel size, the replacement convolutional layer will process images of any size whereas densely connected neural network requires images of particular size. This FCN approach looks at each of the images in the dataset only once and does not go through back propagation like the traditional convolutional neural networks. That is why one of the very popular object detection algorithms is called YOLO (You Only Look Once) [17, 41]. More on YOLO in the methodology section.

Aly [31] discussed about using YOLO (more details in methodology section) for breast masses detection from mammograms. The author mentioned that mammogram worked as an alternative to Magnetic Resonance Imaging (MRI). In MRI, an agent is used for creating a colour contrast in the captured image to which a lot of patients have shown allergic reaction to [42]. Since mammography does not involve use of any allergic agent and is non-invasive, mammography had become the most reliable imaging tool for breast cancer screening. With progression of time, there was a huge collection of breast screening mammograms which increased the workload of radiologists and became more prone to wrong diagnosis. The author proposed a Computer Aided Diagnose (CAD) system with different versions of YOLO to give the radiologists a second opinion about breast cancer diagnosis to help speed up and optimize their processes. The author compared performance of all 3

versions of YOLO models. Finally, the author concluded that YOLO version3 is the best one for object detection. The author achieved 89.4% of accuracy in the INbreast mammograms [43]. Murat [32] brought attention to skin cancer, one of the most widespread types of cancer in the world. The author proposed a 4 steps algorithm to carry on with the project. Deleting the hair on the lesion in the picture using the DullRazor [44] algorithm, detecting lesion from the region of interest using YOLO v3, using GrabCut [45] algorithm to segment the lesion. The algorithm was tested on PH2 and ISBI 2017 dataset. The algorithm achieved 94.90 and 96 percent detection accuracy on the datasets respectively. G. Chen used Adaptive Hybridized Deep Convolutional Neural Network (AHDCNN) to predict chronic kidney disease at an early stage. The author focused on image segmentation in MRI images. The ADHCNN model achieved a F_1 score of 97%.

From the above discussion, it is apparent that there is a definite need for intra-operative cancerous tumor detection technology as from the literatures by A. Chung [27], Y. Wang [29], Aly [31] for prostate cancer detecting using SVM on MRI images, by Brunese [35] for Covid detection using VGG-16 on Chest X-ray, the available processes are not intra-operative.

2.4 Object detection in non-medical field

Some other non-medical application of YOLO was found to be Agricultural Greenhouses Detection in high-Resolution Satellite Images [38], accurate and robust monitoring method of full-bridge traffic load and distribution [39], airplane detection in transfer learning [40]. With the advancement of civilization, agricultural greenhouses play an important role in satisfying the farm products needs for humans. Rapid construction and expansion of agricultural greenhouses have negative impact on the environment such as occupying high-quality cultivated land, damaging soil, polluting the environment with plastic wastes. The adverse effect on the environment caused by greenhouses has made detecting agricultural greenhouses an annual routine work for the Ministry of Natural Resources in China. Images collected from satellite, unmanned aerial vehicles (UAV) along with computer vison have proven useful for greenhouses detection. Identifying traffic load distribution on the bridge is of utmost importance for taking precautions against failure. Existing traffic load monitoring (TLM) are not reliable as they are not able to meet the requirements of real-time, accuracy and lighting robustness simultaneously. The researchers made use of weigh-in-motion systems (WIMs), computer vision to develop a traffic load monitoring system.

2.5 Project content

The images of the Kidney in this research had been extracted from Kidney cancer surgery videos performed by surgeons using da Vinci Xi robot. The images were collected from publicly available partial robotic nephrectomy videos in YouTube. After extracting the regions to be analysed from the

videos, the images were divided in 3 folders. One folder had been kept as back up. Images from one folder had been labelled using LabelImg [46] for to be used in YOLO v4 [47] model. The other folder was used in AlexNet [22], VGG-16 [25] with GRAD-CAM [37] to get a visual explanation of how the neural network was working. The image folder being used for classification had been classified into cancerous tissue, non-cancerous tissue and fatty tissue. The classification step was classifying between cancerous tissue, non-cancerous tissue and fatty tissue.

While H. Murat [32] used NVIDIA GTX1080 Ti GPU, G. Aly [31] used RTX 2080 Ti GPU, M. Li used two Titan RTX GPU, in this project the algorithms had been trained using NVIDIA Quadro P4000 GPU.

The gradient based overfitting method as in [21] had been used to tune the hyperparameters of different network architectures. The evaluation metrics Loss VS epochs, confusion matrix as in [24], [25], [26], [31] had been used to compare the performance of different models with each other. While all the medical applications mentioned above used different ML and DL algorithms to detect cancers from prostate, skin, breast, kidney, they mostly focused on microscopic images in ultrasound or MRI biopsy, this project focused on detecting cancerous tissue on macroscopic level and real-time during surgery.

After training using the 4 algorithms mentioned above, a comparison had been discussed between those to explain why one is better than the other and finally decide on which one would be used for live detection and which one would be used for classification and localization.

3. Methodology

3.1 Introduction

This section describes how the dataset for classification and localization of cancerous tissue and for detecting cancerous tissue real-time during surgery was collected. The dataset was prepared in 6 different ways before the dataset that worked best could be found. In the results section, it had been discussed in detail how 6 different datasets affected the result and why one dataset was preferred compared to the other five datasets. 1^{st}, 3^{rd}, 5^{th} dataset has 3 classes (cancerous, non-cancerous and fatty tissue). 2^{nd}, 4^{th}, 6^{th}, dataset has 2 classes (cancerous and non-cancerous tissue). Then comes the section for describing how the dataset for object detection was prepared.

The section also describes the 4 Computer Vision algorithms used in this project. For the classification and localization tasks, GRAD-CAM was used to get a visualization of where in the image the algorithm was looking at to give a classification result. Intermediate CNN layer activations and filters are observed to help in tuning the hyperparameters.

This section covers the basics of CNN and discusses about the architectural difference of AlexNet, VGG-16 and ResNet. it also covers the basics of FCN and discuss the working principle of YOLOv3 and YOLOv4.

3.2 Collection, preparation and description of 6 dataset

The images for the dataset were collected from publicly available partial robotic nephrectomy videos in YouTube. The images were cropped from the videos, while attempting to exclude the robotic arms from the videos in the cropped images. An example of a cropped photo is shown in Figure 10.

(a) (b)

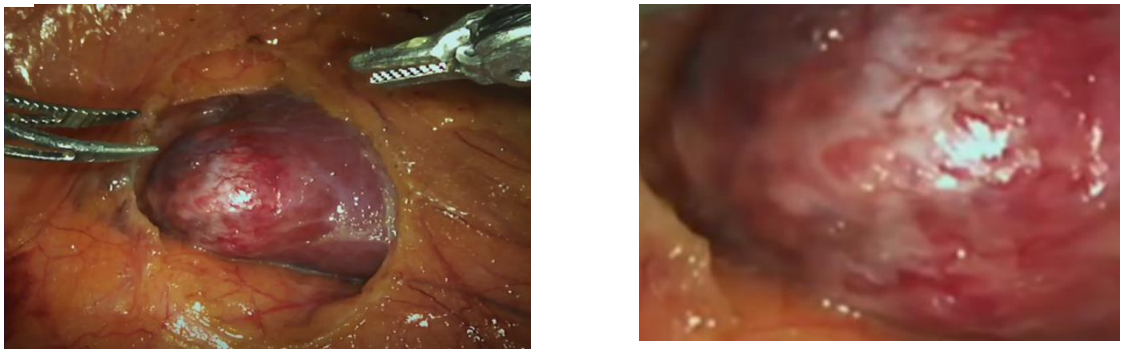


Figure 10: (a) Image cropped from video, (b) Image cropped from (a) [48]

The cropped photos were cropped again to only include either cancerous tissue or non-cancerous tissue or fatty tissue. In Figure 10(b), only the cancerous tissue was included from the entire image. This was done because there was not enough image in the dataset for the algorithms to pick up patterns for cancerous, non-cancerous and fatty tissue from the entire image during training.

The images were collected from references [48-56] and then cropped like Figure 10 (b). The cropped images were stored in their respective folders. For example, if a cropped image contained cancerous tissue, it was stored in the folder named cancerous tissue and if an image contained fatty tissue, it was stored in fatty tissue folder. The images in the folder were numbered starting the indexing from 1. The images were resized to 224 x 224 before being fed into the algorithms. The images were collected online. This is the greatest number of related images that could be collected. Any unnecessary images in the dataset will decrease the performance of the algorithm.

3.2.1 1st dataset

In the dataset, there were 44 cancerous tissues of kidney, 62 non-cancerous tissues (most were kidney, some other healthy organs were also included to increase the number of non-cancerous tissues) and 37 fatty tissue photos. The dataset was divided into training, validation and test set following the pattern described in Table 1 below-

Table 1: Train, validation, test division for 1st dataset

	Total	Train	Validation	Test
Cancerous tissue	44	30	9	5
Non- cancerous tissue	62	40	13	9
Fatty tissue	37	21	10	6

In this case, fatty tissue was imposing problem in getting classified and the highest accuracy was 53%. Most of the output from GRAD-CAMs were also wrong.

3.2.2 2nd dataset

Fatty tissue images were in lower quantity in training, validation and test set and in turn fatty tissue imposed a challenge in correctly classifying. For the 2nd dataset, fatty tissue and non-cancerous tissues were combined and the target was to do binary classification between cancerous and non-cancerous tissue. For the dataset, train, validation and test set division were as follows –

Table 2: Train, validation, test division for 2nd dataset

	Total	Train	Validation	Test
Cancerous tissue	44	30	9	5
Non- cancerous tissue	99	61	23	15

Since the quantity of non-cancerous tissue images were more than double than the cancerous tissue images, all the non-cancerous tissues got correctly classified but the performance for classifying the cancerous tissues were not good. The accuracy was 91%, but it was coming from the greater number of non-cancerous tissue.

3.2.3 3rd dataset

It was noticed that when the images were further cropped from the original images to only include cancerous tissue portion or non-cancerous tissue portion, the images got blurrier as shown in Figure 11(a).

So, for the 3rd dataset, unsharp masking was applied to increase the sharpness of the images as shown in Figure 11(b).

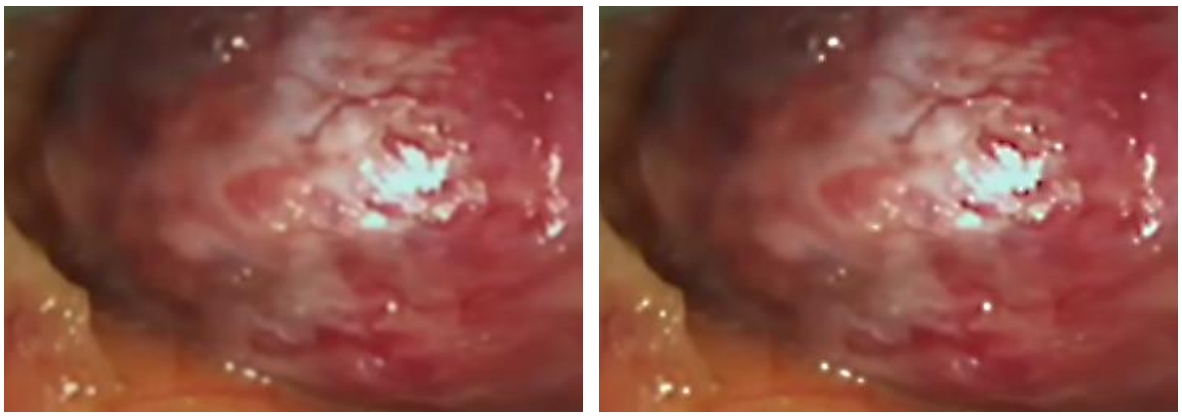


Figure 11: (a) Blurry image (b) Unsharp masking applied

The images from the 1st dataset cancerous tissue were augmented. The images were mirrored, rotated 90^0 clockwise, rotated 180^0 clockwise, rotated 45^0 clockwise. So, 1 train cancerous tissue image was made into 5 images. The train, validation and test division were as follows-

Table 3: Train, validation, test division for 3rd dataset

	Total	Train	Validation	Test
Cancerous tissue	164	150	9	5
Non- cancerous tissue	62	40	13	9
Fatty tissue	37	21	10	6

3.2.4 4th dataset

This dataset was made almost the same way as the 3rd dataset. The only difference here was the number of classes here were 2. The train test split was as follows-

Table 4: Train, validation, test division for 4th dataset

	Total	Train	Validation	Test
Cancerous tissue	164	150	9	5
Non- cancerous tissue	99	61	23	15

3.2.5 5th dataset

It was ensured that training cancerous, non-cancerous and fatty tissue had equal number of images. Unsharp masking was no longer considered as an option here. The divisions were as follows-

Table 5: Train, validation, test division for 5th dataset

	Total	Train	Validation	Test
Cancerous tissue	119	105	9	5
Non- cancerous tissue	127	105	13	9
Fatty tissue	121	105	10	6

3.2.6 6th dataset

The data preparation step is almost the same as 5th dataset. The only difference here was that there were 2 classes. The target was to make equal number of cancerous and non-cancerous tissue for the training set. Train, validation and test set division-

Table 6: Train, validation, test division for 6th dataset

	Total	Train	Validation	Test
Cancerous tissue	164	150	9	5
Non- cancerous tissue	188	150	23	15

3.2.7 Object Detection Dataset

For object detection, the images were left as in Figure 10 (a) without further cropping unlike classification and localization. LabelImg [46] is an open source graphical image annotation tool. For Object detection training there were 56 images in total (49 for training and 7 for testing). The training images were loaded into LabelImg and manually labelled for Cancerous tissue, Non cancerous tissue and Fatty tissue.

3.3 Fundamentals of Convolutional Neural Networks

Images are represented in pixel values in computer vision. A grayscale image has only 1 color channel whereas a color image has 3 color channels.

(a)

-1	-1	-1	-1
0	0	0	0
1	1	1	1
-1	-1	-1	-1

(b)

(c)

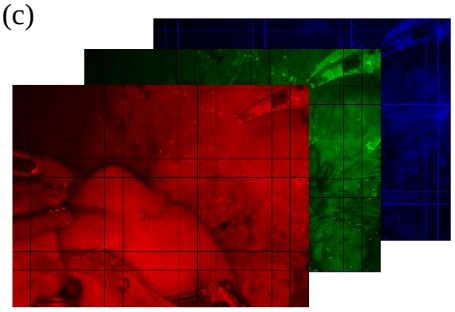


Figure 12: (a) Grayscale image (b) Color image (c) Color image divided into color channels

In Figure 12(a), the grayscale image's pixel intensity values are represented with numbers where 1 is white, -1 is black, 0 is grey. Figure 12(b) shows a color image of kidney tumor which is divided into

its RGB channels in Figure 12(c). In convolution operation, the pixel intensity values of images are multiplied by filters (filters have to be smaller than the images) and the images keep on getting smaller and smaller. For example, an image can be 224 × 224 but the filter will have to be smaller such as 11 x 11. Multiplying the 224 × 224 image by 11 × 11 filter results in (224-11 + 1) × (224-11+1) = 214 × 214 image. Figure 13 shows one example of multiplying a color image divided into RGB color channel into 11 × 11 × 3 filter which results in 214 × 214 image output.

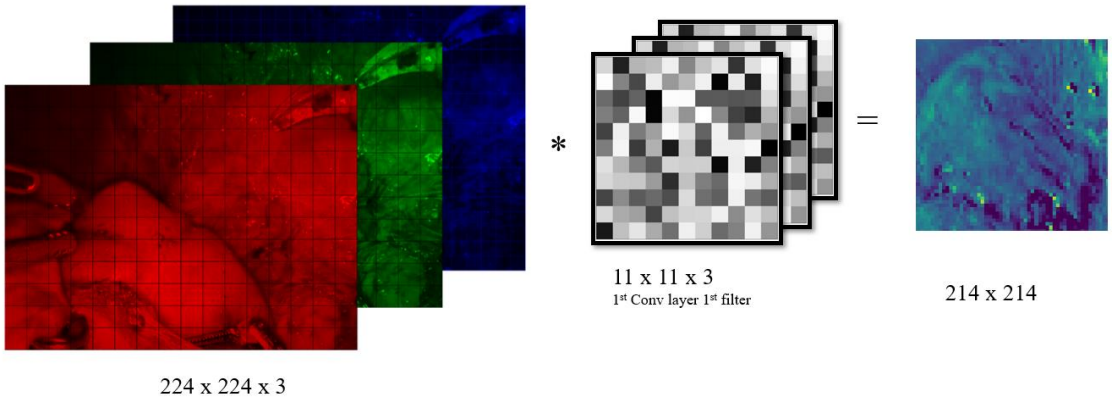


Figure 13: AlexNet 1st layer convolution operation

Convolution layers use a combination of convolution operation, activation transformation, max pooling to shrink images to reduce the number of parameters to process for the neural network.

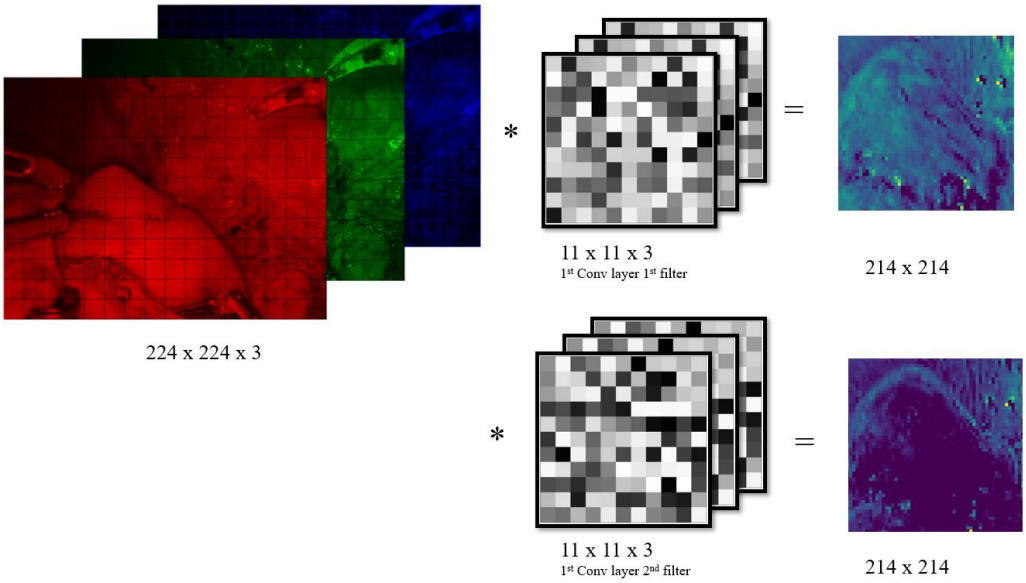


Figure 14: AlexNet 1st layer 2 filters convolution

In Figure 14, multiplying the same image with a second filter results in another 214 × 214 image. Now, Getting the two outputs together as in Figure 15 results in 214 × 214 × 2.

So, if there are n filters in the 1st layer of the convolutional neural network it results in 214 × 214 × n sized image as the output of the 1st layer.

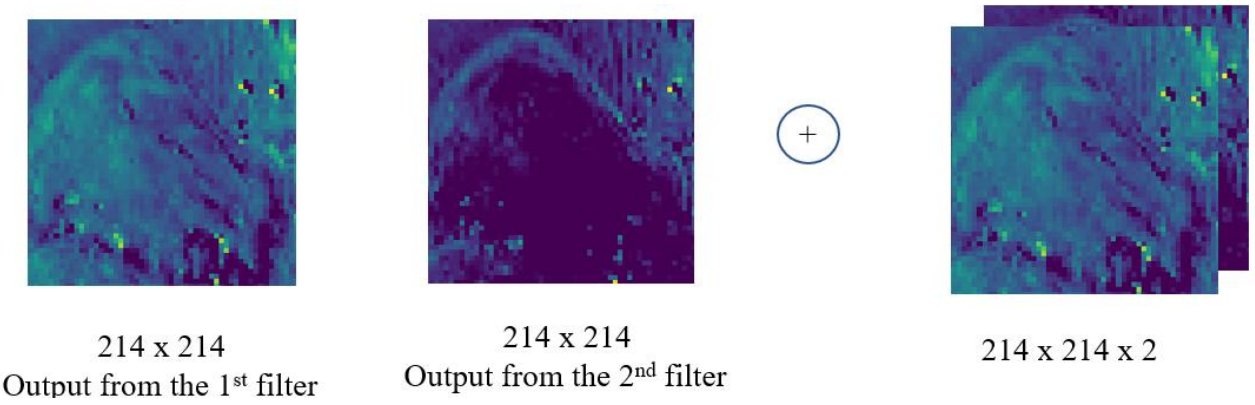


Figure 15: Concatenating output from 1st and 2nd filter in 1st Conv layer

3.4 Information on Max Pooling

Max Pooling is applied to further reduce the number of parameters to process. This enables the algorithm to only focus on the portions of the image where the maximum pixel intensity values are present and delete the portions with lower pixel intensity values.

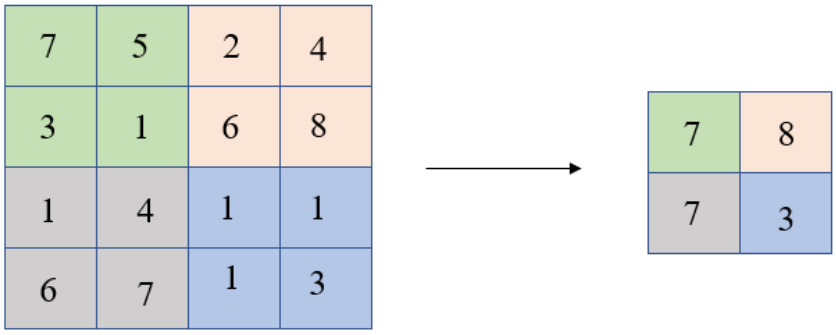


Figure 16: Max Pooling operation

Using Convolution operation, activation transformation, max pooling, the AlexNet changes the size of the input image from 224 × 224 × 3 as show in Table 7 -

Table 7: Simple AlexNet network parameters

Layer (type)	**Output shape**	**Param #**
conv2d (Conv2D)	(None, 54, 54, 96)	34944
max_pooling2d (MaxPooling2D)	(None, 26, 26, 96)	0
conv2d_1 (Conv2D)	(None, 26, 26, 256)	614656
max_pooling2d_1 (MaxPooling2D)	(None, 12, 12, 256)	0
conv2d_2 (Conv2D)	(None, 12, 12, 384)	885120
conv2d_3 (Conv2D)	(None, 12, 12, 384)	1327488
conv2d_4 (Conv2D)	(None, 12, 12, 256)	884992
flatten (Flatten)	(None, 25600)	0
dense (Dense)	(None, 4096)	104861696
dense_1 (Dense)	(None, 4096)	16781312
dense_2 (Dense)	(None, 5)	20485
Total params: 171,539,843		
Trainable params: 171,539,843		
Non-trainable params: 0		

It can be seen that with each convolutional layer, the image is getting smaller (which results in less parameters to process) and the channel depth is getting larger. Each of the channels capture different features of the image starting from the edges and gradually building up to the entire image.

Convolutional Neural Networks can learn local patterns in an image as compared to Densely Connected Neural Networks which learn global patterns. If a CNN learns a pattern in any part of the image it will recognise that pattern in any image. This results in the computations becoming less expensive and faster. In the Figure 12(b), the image is 224 × 224 and has 3 color channels. Using the image in a Densely Connected Neural Network results in 224 × 224 × 3 = 150,528 input features.

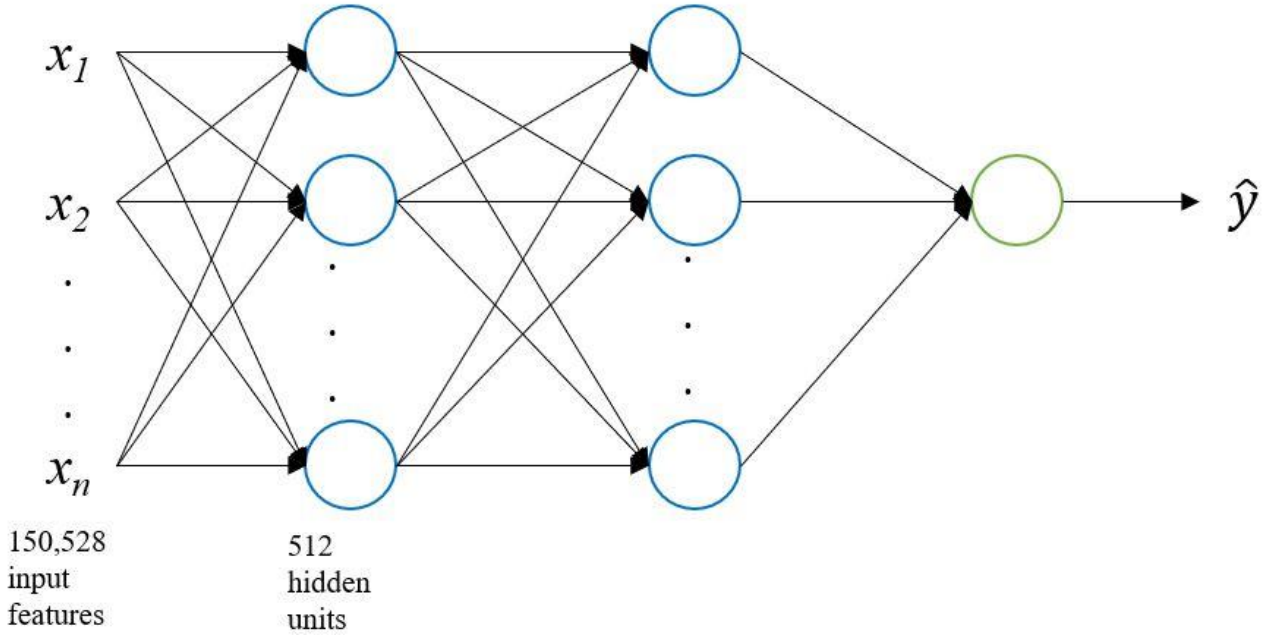


Figure 17: Densely connected Neural Network. *X* are input features; blue circles are hidden units and green circle is the softmax output.

If the input layer is connected to a hidden layer with 512 hidden units as show in Figure 17, then it results in 150,528 × 512 = 77,070,336 parameters just in the 1st hidden layer. The properties of (1) learning patterns of translation invariance and (2) learning spatial features in the images makes CNN suitable for classification and detection on images.

3.5 Convolutional Neural Network architectures

Stacking convolutional neural networks on top of each other with Rectified Linear Unit activation, followed by pooling layer and again convolutional layers with ReLU activations, a CNN architecture is made. From fundamentals of CNN, max pooling, as the depth of the network increases the smaller the image gets, but the channel depth keeps on increasing as the number of low-level features which are captured by the first few layers is low. With time, different variants of CNN architecture were published. This section covers the network architecture of AlexNet, VGG-16 and ResNet-50.

3.5.1 AlexNet

The architecture for the 2012 ILSVRC challenge winner is known as AlexNet. In Figure 19, the architecture for AlexNet is shown. The grey cubes represent the filters used in the hidden layers as shown in Figure 13 and 14. Since the filters are 3 dimensional (height x width x depth) they are represented as cubes. Bigger cubes such as 54 x 54 x 96 represent bigger height and width. Depth of

the cubes represent the depth of the filters. For example, the cube representing 96 filters is shallow compared to filter 12 x 12 x 256.

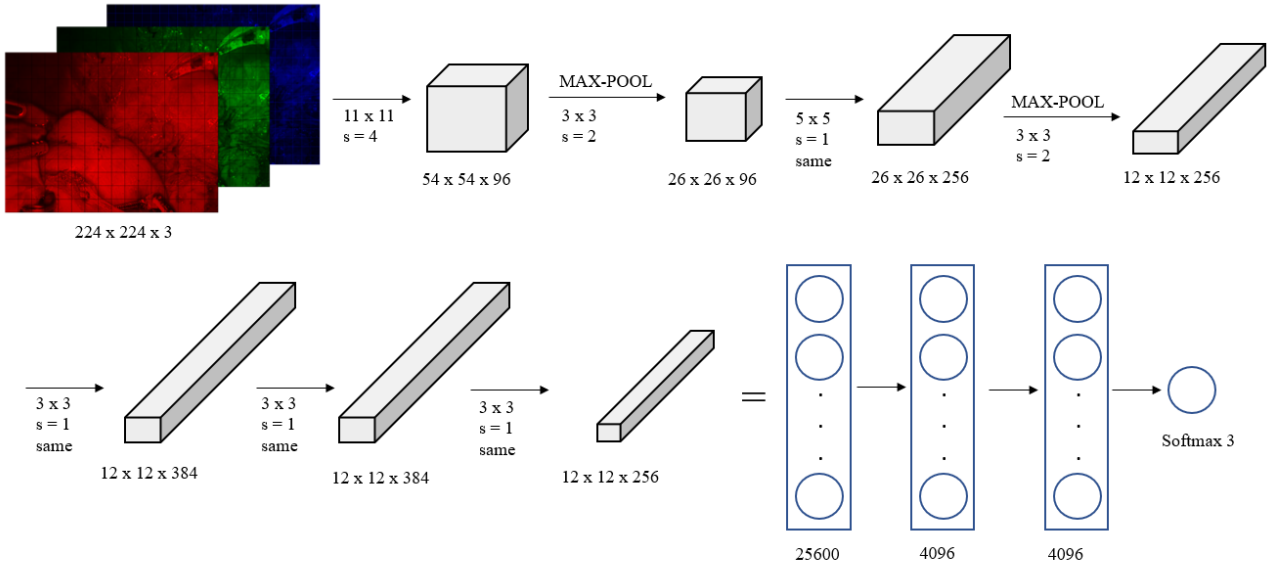


Figure 18: Simple AlexNet architecture

From Figure 18, the network uses 5 convolutional layers with ReLU activation, with the first 2 layers followed by a max pooling layer. The max pooling layer uses 3 × 3 window with stride 2.

The first convolutional layer uses stride 4 but the rest of the convolutional layers use single stride. 2nd, 3rd, 4th and 5th convolution layer use same padding. After the last convolutional layer, the features are flattened (made into a 1-dimensional vector) and then goes through 2 densely connected neural networks each of 4096 hidden units. Then finally in the output layer softmax activation gives prediction for either cancerous tissue or non-cancerous tissue or fatty tissue.

3.5.2 VGG-16

VGG-16 uses 3 × 3 Convolutional filters with stride 1 and same padding, instead of trying out different hyperparameters for the network architecture. In max pooling, VGG-16 uses 2 × 2 window with stride 2 as shown in Figure 19. This network achieves better performance than AlexNet, as it has been found in [24] that deeper network performs better than shallow nets provided caution has been practiced to make sure the network does not overfit.

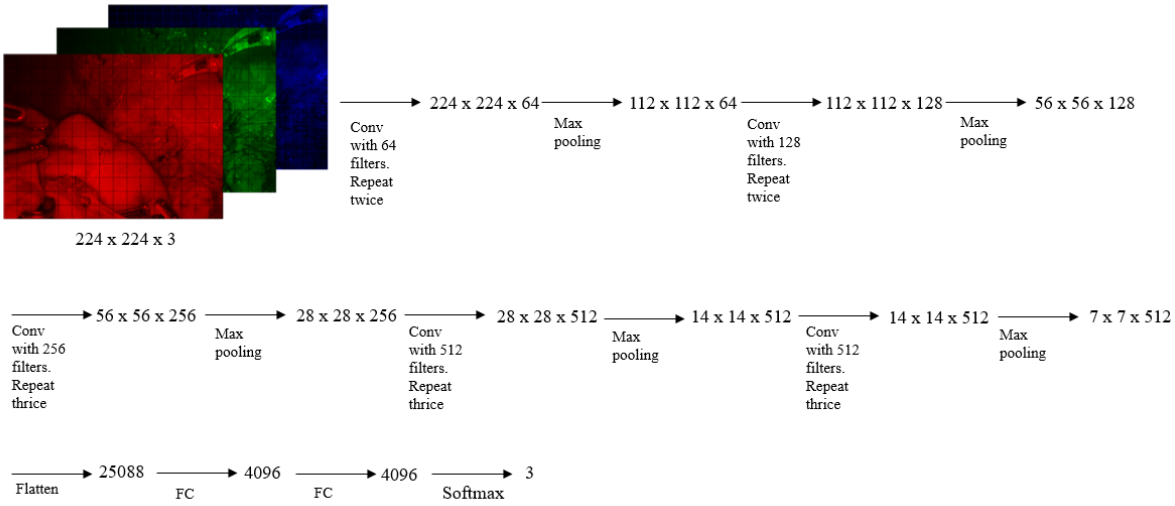


Figure 19: VGG-16 Network architecture

3.5.3 ResNet - 50

Increasing the number of hidden layers in the network, makes the input image smaller and smaller and the network tends to overfit the training set.

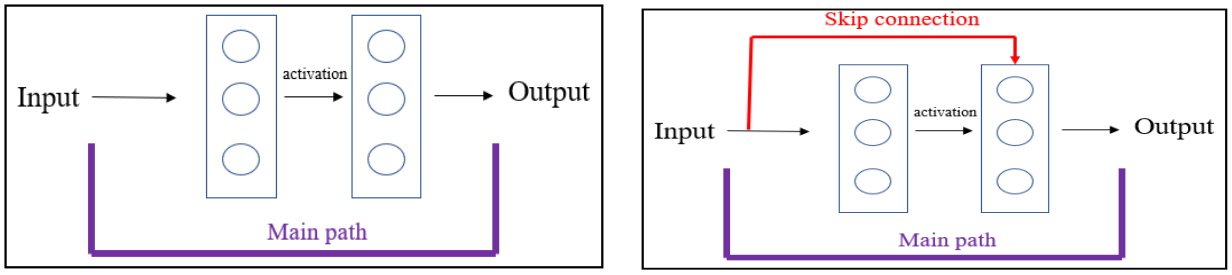


Figure 20: (a) Plain network (b) Skip connection

In Figure 20 (a), this is the plain network that was shown in the previous networks. In Figure 20(b), skip connection is introduced that takes the input and connect that to a layer further into the network. Without going into technical jargon, it is possible to show that the output can be equal to the input by using skip connection. This way, if the hidden units cannot improve the performance, the input is kept as it is without being affected by the hidden units. If the hidden units can improve the performance, then the performance is enhanced by using skip connection. The block in Figure 20(b) is called residual block. ResNet - 50 is made by repeating the residual block 50 times.

3.6 Fully Convolutional Network

Referring to Figure 19, the 7 × 7 × 512 image had been flattened to 25088 dimensional vector for feeding into densely connected neural network. Fully Convolutional Network (FCN) suggests using a 1 × 1 convolution to preserve spatial information instead of flattening the matrix into a vector. For example,

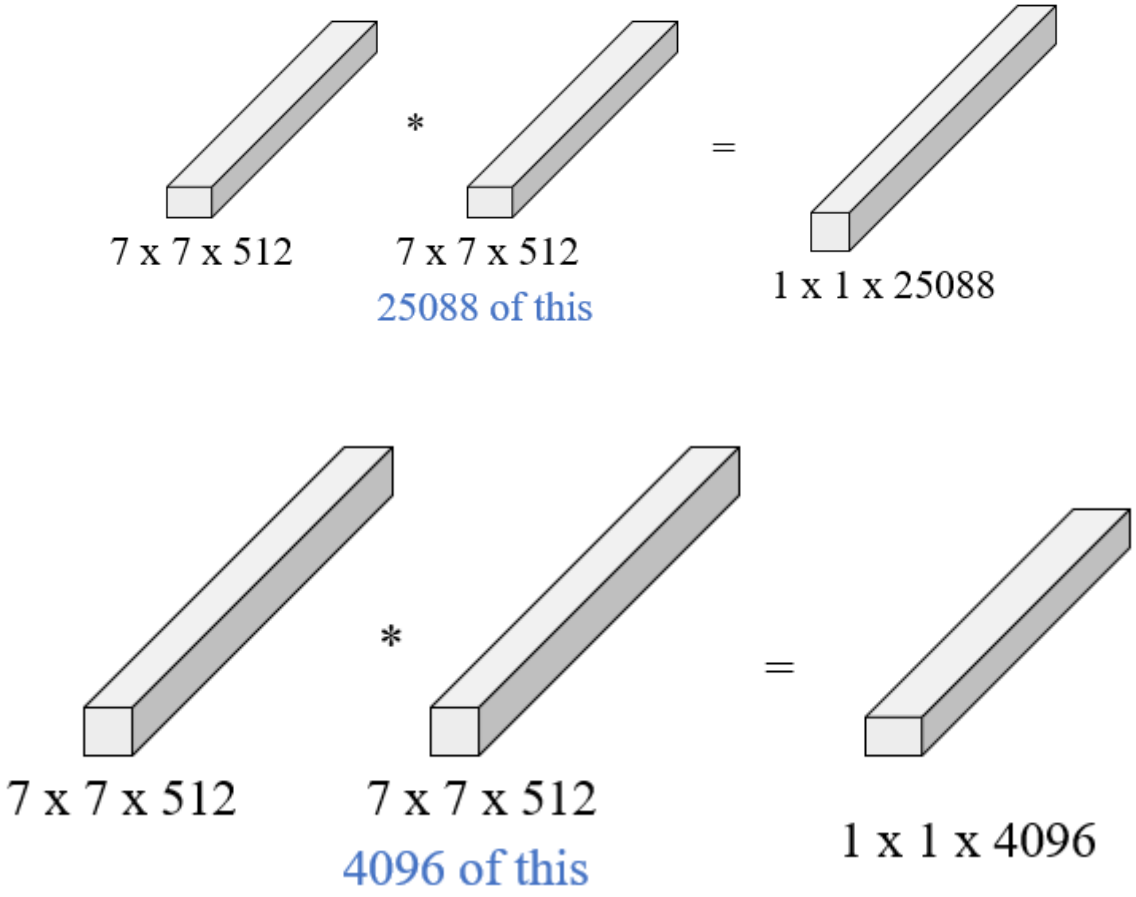


Figure 21: 1 x 1 convolution conserving spatial information

In Figure 21, the flattened layer and the first dense layer had been converted to convolutional layer. In the same way the rest of the dense layers can be converted to convolutional layers.

The algorithms mentioned before such as AlexNet, VGG-16, ResNet-50 are used for classification and localization, where only one object from the image is classified and localized. In object detection (YOLOv3, YOLOv4), multiple objects are classified from the image and boundary boxes are drawn around the object to localize those. Previously, where the output was just predicting the class, here the output is at least 5 dimensional vector.

The 5 dimensions being the confidence score P_c (the probability of there being an object in the bounding box), coordinates for the bounding boxes (X_{min}, Y_{min}, X_{max}, Y_{max}), depending on the number

of classes C_n there are more dimensions in the output. If there are 100 classes C_{100}, then the output will be a 105 dimensional vector.

Now, referring back to Figure 12(b), assuming after going through a FCN, the image was divided into 12 × 12 grid cells as in Figure 22 (a). The output is 12 × 12 × 7 in shape. Each of the grid cell's depth is 7 because [P_c, X_{min}, Y_{min}, X_{max}, Y_{max}, C_2] (C_2 representing 2 classes (Cancerous tissue and non-cancerous tissue)).

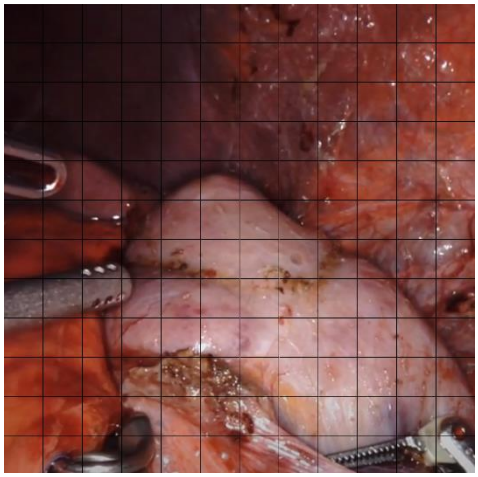
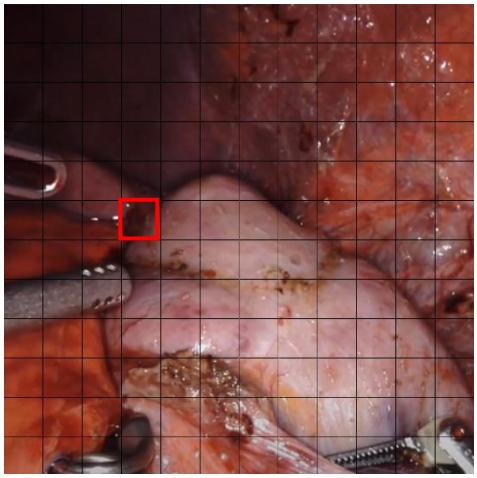
(a)  (b) 

Figure 22: (a) Output from FCN (b) Red grid cell containing information about the tumor

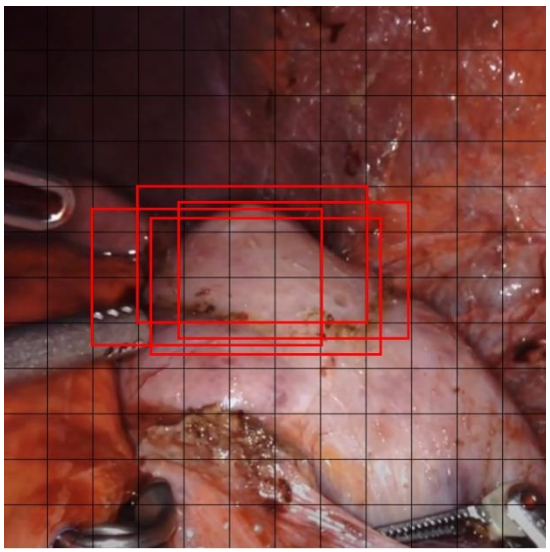
When sliding the convolutional window it gives output as -

[1, X_{min}, Y_{min}, X_{max}, Y_{max}, 1, 0]. The 1 being the confidence score that the grid cell contains an object. The next 4 numbers for the bounding box coordinates, the 1 after is the score for finding cancerous tissue class, the 0 is for non-cancerous tissue class. Similarly, all the grid cells contain the 7 dimensional vector and this is how the spatial information in FCN is conserved unlike Convolutional Neural Networks discussed before.

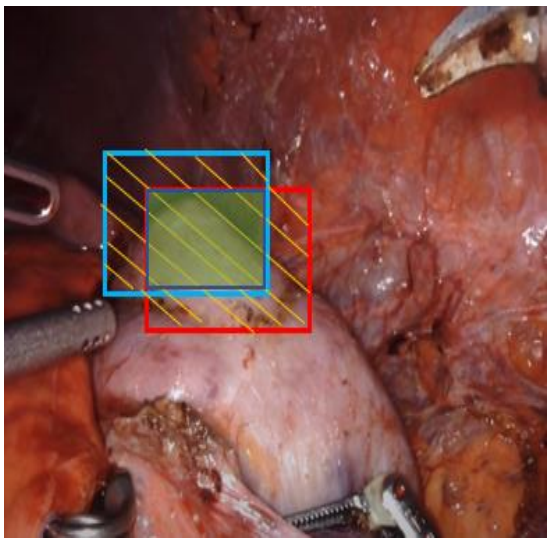
3.7 YOLO v4

YOLOv4 builds up on the idea of FCN. YOLO stands for You Only look Once. All the grid cells in Figure 22, contains spatial information about the objects those contain. So, the grid cells around the tumor in Figure 22 will try and make bounding boxes for the tumor and will look like Figure 23 (a).

(a)



(b)



(c)

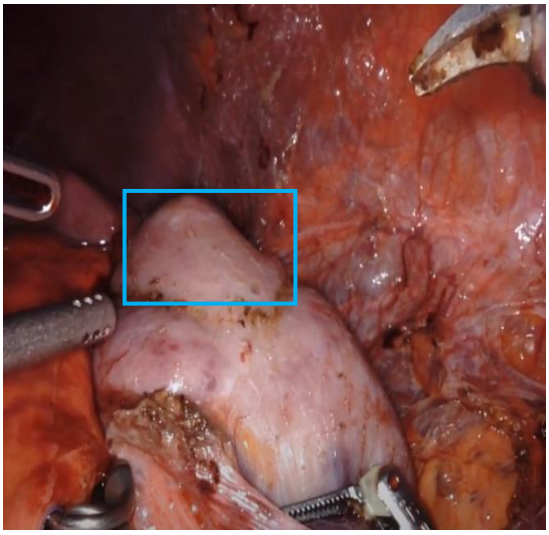


Figure 23 (a): Grid cells around the tumor predicting bounding boxes (b) Intersection over Union (c) Non-max suppression outputting 1 bounding box per object

Each of these bounding boxes has a confidence score (P_c) coming from the grid cell that is responsible for making the bounding box. The steps for YOLO are explained below-

1. All the boxes with P_c <= 0.6 are discarded.

2. With the remaining boxes, the boxes with largest P_c output is taken as a prediction. Then, the concept of Intersection over Union (IoU) (Figure 23(b)) is used.

In Figure 23(b), the red bounding box is the ground truth. The blue bounding box is the predicted bounding box. The green 50% opacity square is the area of intersection and the orange lines represent the area of union.

Intersection over Union = $\frac{Area\ of\ intersection}{Area\ of\ union}$

A perfect IoU is 1. The higher the IoU, the more accurate the bounding box prediction is.

3. The remaining boxes with IoU >= 0.5 with the box output in the step 2 is discarded.

In the end there is only 1 bounding box for each object. This technique of removing all the redundant bounding boxes is called non-max suppression as in Figure 23(c).

4. Performance Analysis

4.1 Introduction

This section discusses about the results from the output of object detection and classification/localization algorithms. The aim is to first discuss the results of object detection algorithm and then classification and localization algorithms. The results are presented in this order to give an indication of how this can be used in real life surgery. The object detection algorithm detected cancerous tissue in a global range. After the identified tumors had been cut off, the classification and localization algorithm could be used to detect tumors in close range.

4.2 Object Detection Result

The object detection system which was used to detect tumor on a global range inside the patient during surgery was conducted on Intel Core i7 10^{th} gen, equipped with NVIDIA Quadro P4000 GPU and 32 GB of RAM. The dataset had 56 images in total. 49 images were used for training and 7 images were left for testing. Python based open-source image annotation tool LabelImg was used for preparing the dataset for training [46]. The YOLOv4 algorithm used for object detection was written by AlexeyAB [57]. The open-source code was downloaded from GitHub and using OpenCV as the vision engine the images were loaded into the model. The training was run for 15 hours with image augmentation activated. After training was done, the algorithms were tested with the images from the test set and on the videos from which the test set images had been extracted.

Before training the algorithm on the windows PC, it was implemented in Google Colab virtual machine using YOLOv3. It was done because implementing YOLO on a virtual machine is comparatively easier and to get a quick indication of the performance. The evaluation metrics that were used are as follows-

- Precision, recall
- mean Average Precision (mAP)
- Frames Per Second (FPS) (For video data)

4.2.1 Evaluation with Metrics

OpenCV on virtual machines do not have the capability of opening the detected image or video file instantly after running detection. YOLOv3 was not capable of running detection on videos on a virtual machine. Comparing results of YOLOv3 on the virtual machine with YOLOv4 on windows machine-

Table 8: Result comparison

Detection algorithm	**Precision**	**Recall**	**mean Average precision**	**Frames per second**
YOLOv3 on virtual machine	0.88	0.62	0.758	Not applicable
YOLOv4 on windows	0.98	0.99	0.974	21.4

From Table 8 the mAP of YOLOv4 on windows is better than YOLOv3 on virtual machine. Also, it was not possible to run detection on videos on the virtual machine. Therefore, in terms of evaluation metric, the YOLOv4 is better for tumor detection.

4.2.2 Visual evaluation

The training on the virtual machine and on the windows machine was run for 6000 iterations. The weights were saved every 1000 iterations for each case.

The results were tested on the test set images on the virtual machine. It was found that for some of the images the model was overtrained (performance decreased) at 6000 iteration and for some images 6000 iteration was the optimal one. It was found that for each of the test image it was required to test with 3000 iteration weights, 4000 iteration weights, 5000 iteration weights to find out which weights worked best for a particular test image.

In Figure 24, some of the detected test images with their optimum set of weights are mentioned.

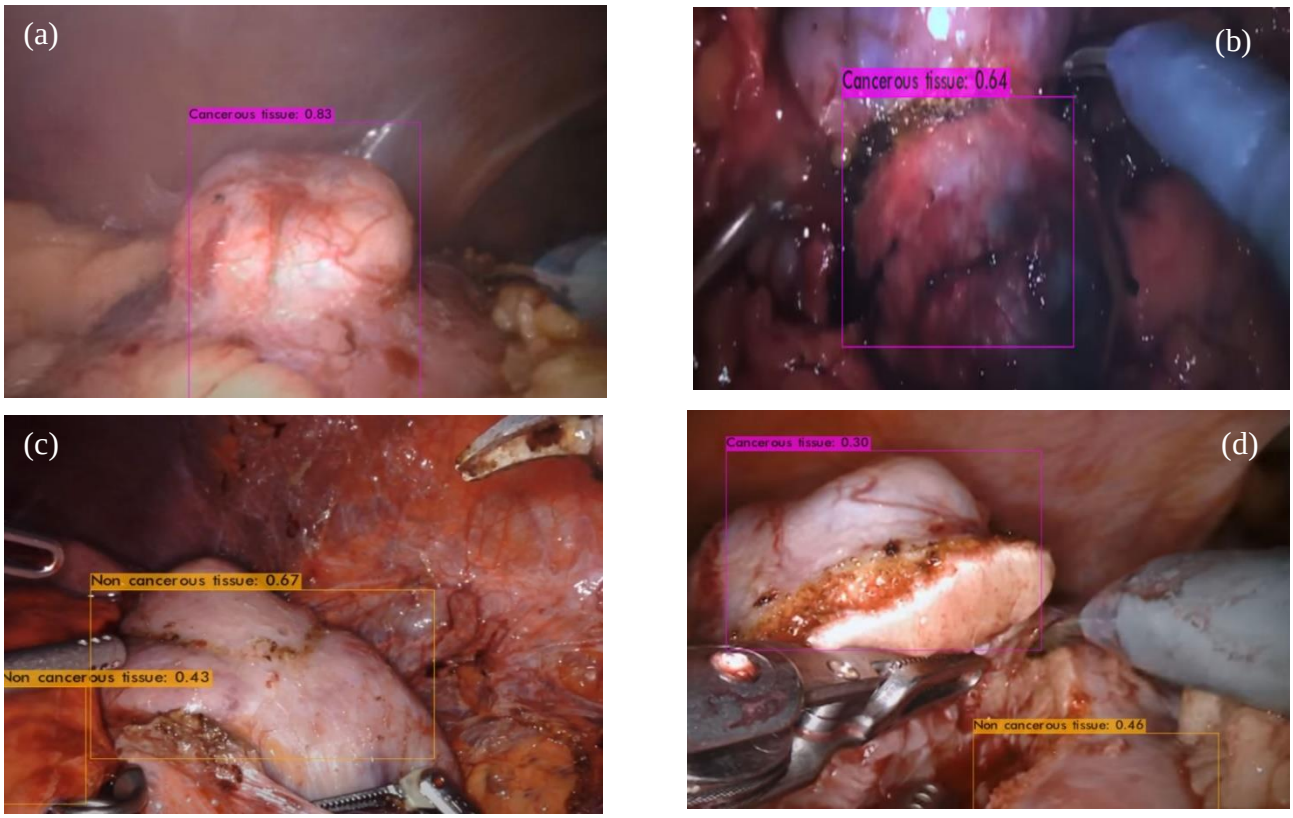


Figure 24: Tumor detection on virtual machine on still images using different weights at different iterations. (a) 4000 iterations (b) 5000 iterations (c) 3000 iterations (d) 4000 iterations

From Figure 24, it is apparent that for YOLOv3 on a virtual machine to work, different set of weights such as 3000 iteration weights, 4000 iteration weights, 5000 iteration weights, 6000 iteration weights will have to be attempted on the test image to identify the optimum set of weights for that image and then identify tumors from there.

On the other hand, YOLOv4 on windows was run on videos and the results were extraordinary. The algorithm was able to detect tumors from real time surgical videos with more than 90% confidence. Some of the examples are shown in Figure 25. Please refer to this video link for a visualization

https://www.dropbox.com/sh/42dy79r2wyjrsq3/AAASkoWs26bFjFkVxkGJfOSwa?dl=0

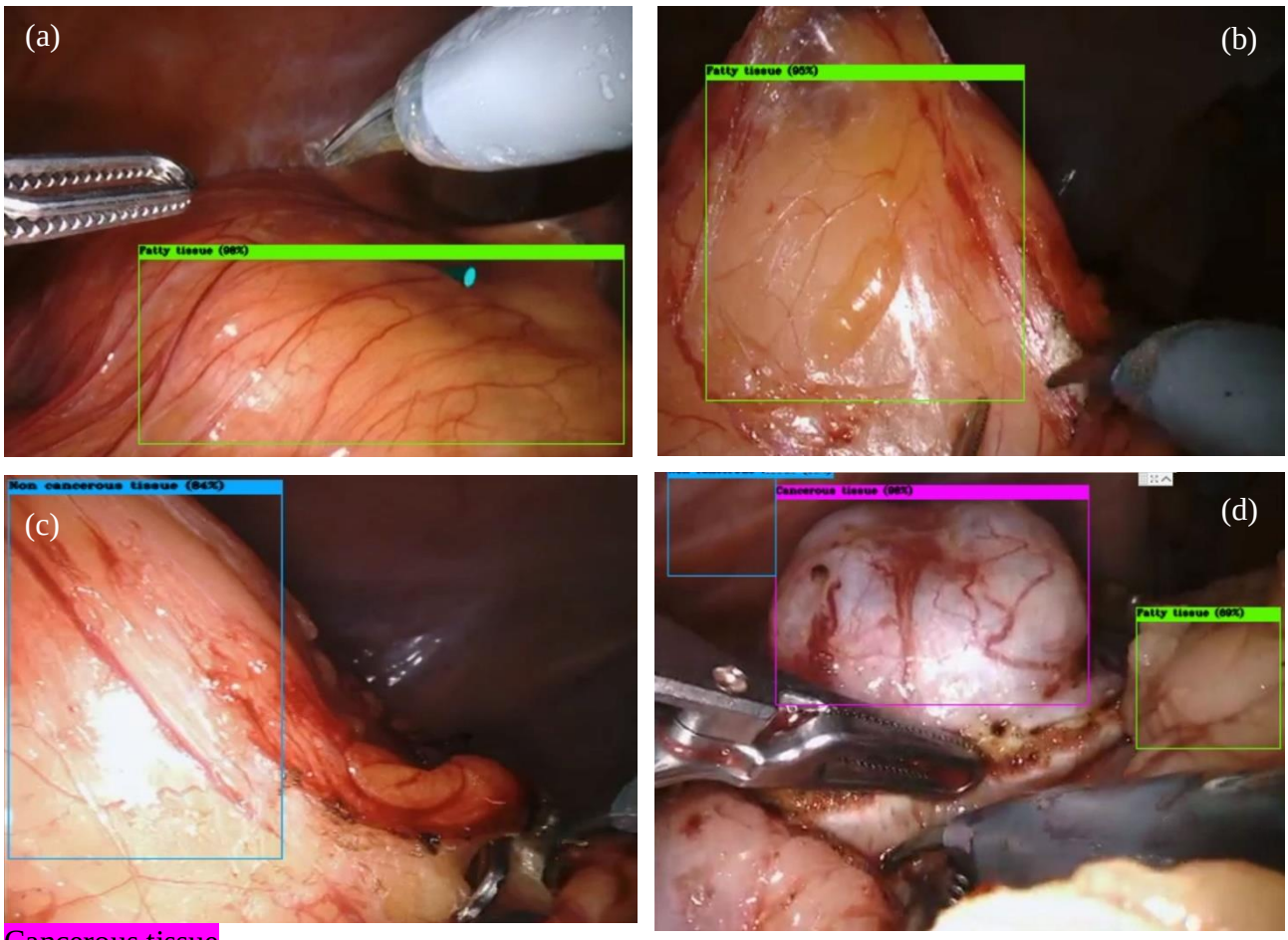


Figure 25: Tumor detection on windows machine on videos (a) fatty tissue (b) fatty tissue (c) Non-cancerous tissue (d) Cancerous tissue, Non cancerous tissue, fatty tissue.

From Figure 24 and 25, the YOLOv4 algorithm on videos does not need adjustment of weights unlike YOLOv3 on virtual machine. On both cases, the pink box is for cancerous tissue, green is for fatty tissue on a windows machine, fatty tissue did not get detected on the virtual machine, Non cancerous tissue is blue on windows machine and yellow on virtual machine.

So, for global range tumor detection real-time in the patient, YOLOv4 on windows machine is good for detection on videos and YOLOv3 on virtual machine is good for detecting on still images. But for the virtual machine, 4 set of weights will have to be attempted before a decision can be made about one image.

4.3 Classification and Localization Result

The dataset was prepared in 6 different ways for classification and localization. Here, in depth discussion is done on the 1st dataset with AlexNet, 5th and 6th dataset with variations of VGG-16. Why the other 3 datasets did not perform well was briefly explained in the methods section. ResNet50 was applied to introduce skip connection to tackle overfitting in a deeper network, but the network was still overfitting. So, discussion is limited to variation of AlexNet and VGG-16.

The performance was evaluated using the following evaluation matrices-

- Loss VS epochs curve
- Confusion Matrix
- Accuracy
- Gradient based Class Activation Mapping (GRAD-CAM) [37]
- Class Activations (Feature maps)

4.3.1 Loss Vs epochs curve

AlexNet with 7 × 7 window filters in the 1st layer with stride 4 and augmentation give the plot as shown in Figure 26 (a).

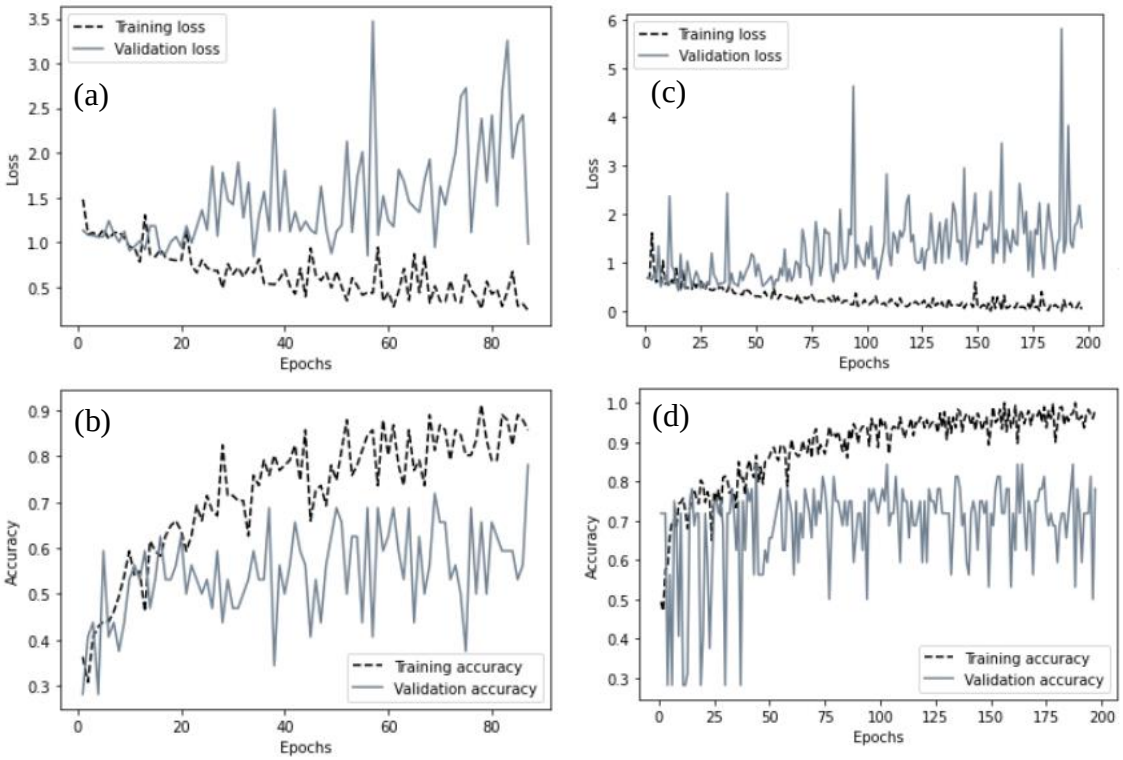


Figure 26: (a) Loss VS epoch of AlexNet (b) Accuracy VS epoch of AlexNet

(c) Loss VS epoch of VGG-16 ($lr=10^{-4}$) (d) Accuracy VS epoch of VGG-16 ($lr=10^{-4}$)

Figure 26 (a) gives an indication of overfitting [58] by the validation loss getting higher with increasing iteration and the validation loss fluctuating a lot. The model overfitted in just 87 iterations. A deeper network was needed to get better performance. That is why VGG-16 was employed with 16 layers compared to 7 layers in AlexNet.

The validation loss became a bit closer to training loss in Figure 27(b) but there is still a lot of fluctuation. To reduce overfitting, the learning rate was decreased from 10^{-4} to 10^{-6} for 2 class classification (6^{th} dataset) and to 10^{-5} for 3 class classification (5^{th} dataset) as in Figure 27.

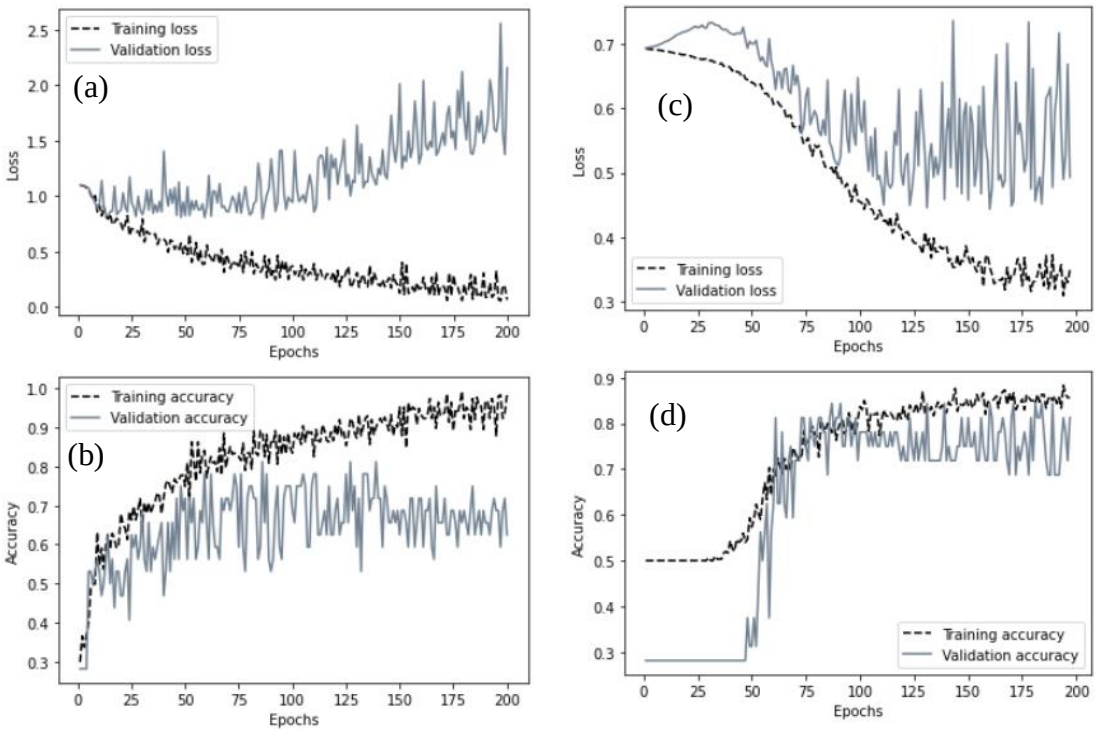


Figure 27: (a) Loss VS epoch of VGG-16 with 5^{th} dataset ($lr=10^{-5}$) (b) Accuracy VS epoch VGG-16 with 5^{th} dataset ($lr=10^{-5}$)(c) Loss VS epoch of VGG-16 with 6^{th} dataset ($lr=10^{-6}$) (d) Accuracy VS epoch VGG-16 with 6^{th} dataset ($lr=10^{-6}$)

Callback, dropout, dropout with callback have also been applied to stop iteration when there are no more improvements. Loss VS Epoch & Accuracy VS epoch curves for those are shown in appendix A.

From appendix A, VGG-16 lower lr callback, lower lr dropout 0.5 and lower lr Dropout 0.5 callback are very close for both the 5^{th} dataset (A1, A2) and 6^{th} dataset (A3, A4). For that, Confusion matrix had been investigated.

4.3.2 Confusion matrix

The confusion matrix for the best performing models for 5th dataset and 6th dataset is shown here. For the rest of the confusion matrix refer to appendix B.

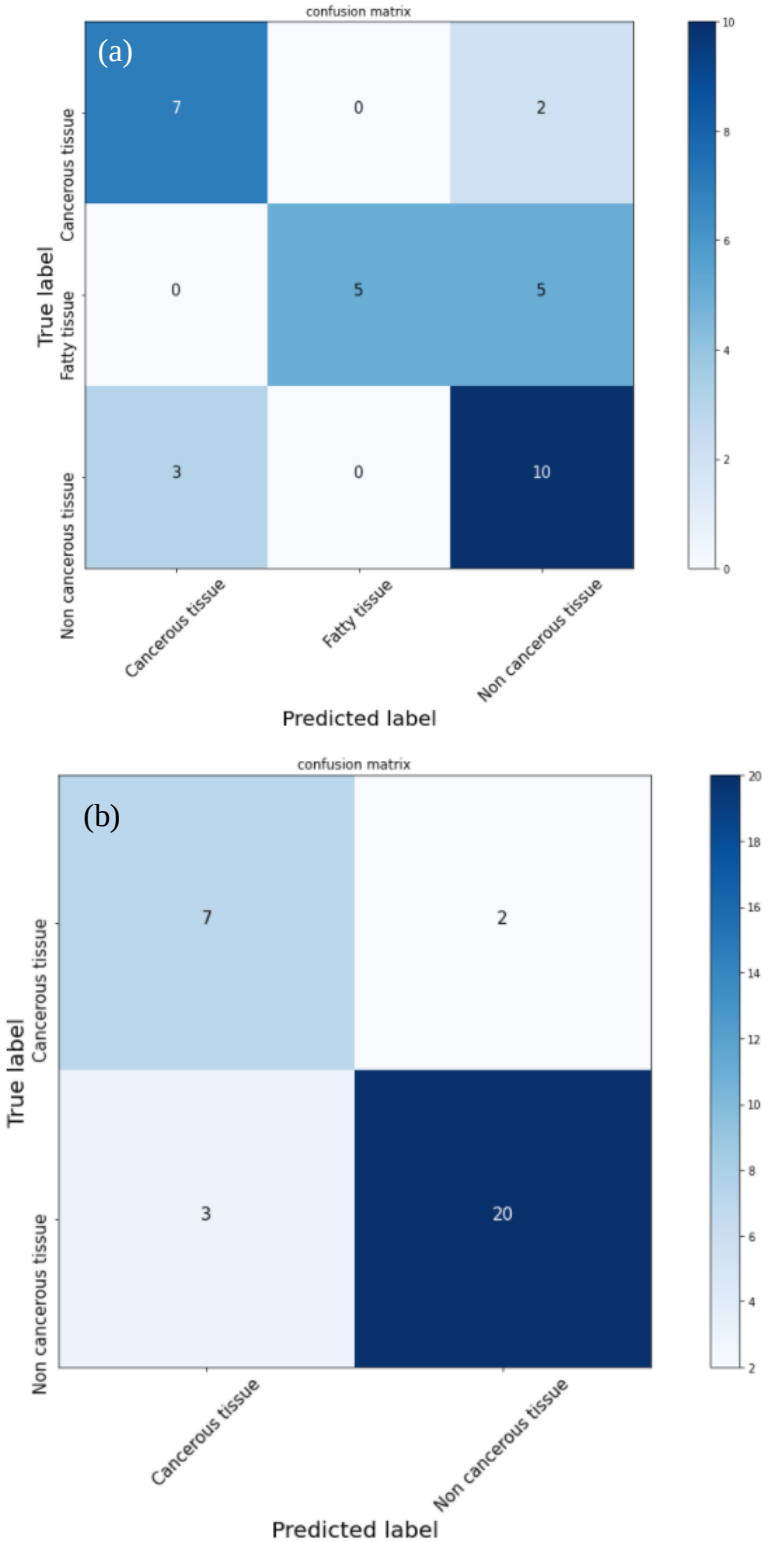


Figure 28: (a) VGG-16 5th dataset, lower lr, Dropout 0.5, callback (b) VGG-16 6th dataset, lower lr callback

Figure 28 (a) and (b) shows that in the y axis is the true label, in the x axis is the predicted label. Out of 9 cancerous tissue, 7 were correctly classified for both cases. In the appendix B, other models with poor performing Loss VS epochs curve was not shown because of their poor performing confusion matrix.

Looking into appendix B (B1, B2), comparison cannot be done between VGG16 5th dataset lower lr callback and VGG16 5th dataset lower lr Dropout 0.5 callback for 3 class classification, between VGG-16 6th dataset 2 lower lr using callback and VGG-16 6th dataset 2 lower lr dropout 0.5 callback.

For getting a proper comparison of performance between these two, accuracy, GRAD-CAM, Class Activations had been used.

4.3.3 Accuracy

Accuracy is not a good comparison factor since the 2 models to compare have the same accuracy. For more information refer to Table 9.

Table 9: Showing that the accuracy are the same for the models to compare

Model	Accuracy
5th data lower lr callback	69%
5th dataset lower lr Dropout 0.5 callback	69%
6th dataset lower lr callback	84%
6th dataset lower lr dropout 0.5 callback	84%

4.3.4 Gradient based Class Activation Mapping (GRAD-CAM)

There are 5 cancerous tissue images in the test set. Here, it is tested which algorithm can give correct prediction for the cancerous tissue and also highlight the cancerous tissue portion on the image. The networks output 2 or 3 probabilities for each image depending on whether 2 (6th dataset) or 3 (5th dataset) class classification is being done. The highest probability region gets highlighted as red in the image. For example, the Figure shown in Figure 29 gets 3 probability output as [0.971,0.00035,0.028] when 3 class classification is done. The first probability is for cancerous tissue, the second number is for fatty tissue, the third is for non-cancerous tissue. Since the cancerous tissue probability is high, that gets highlighted with red in the image.

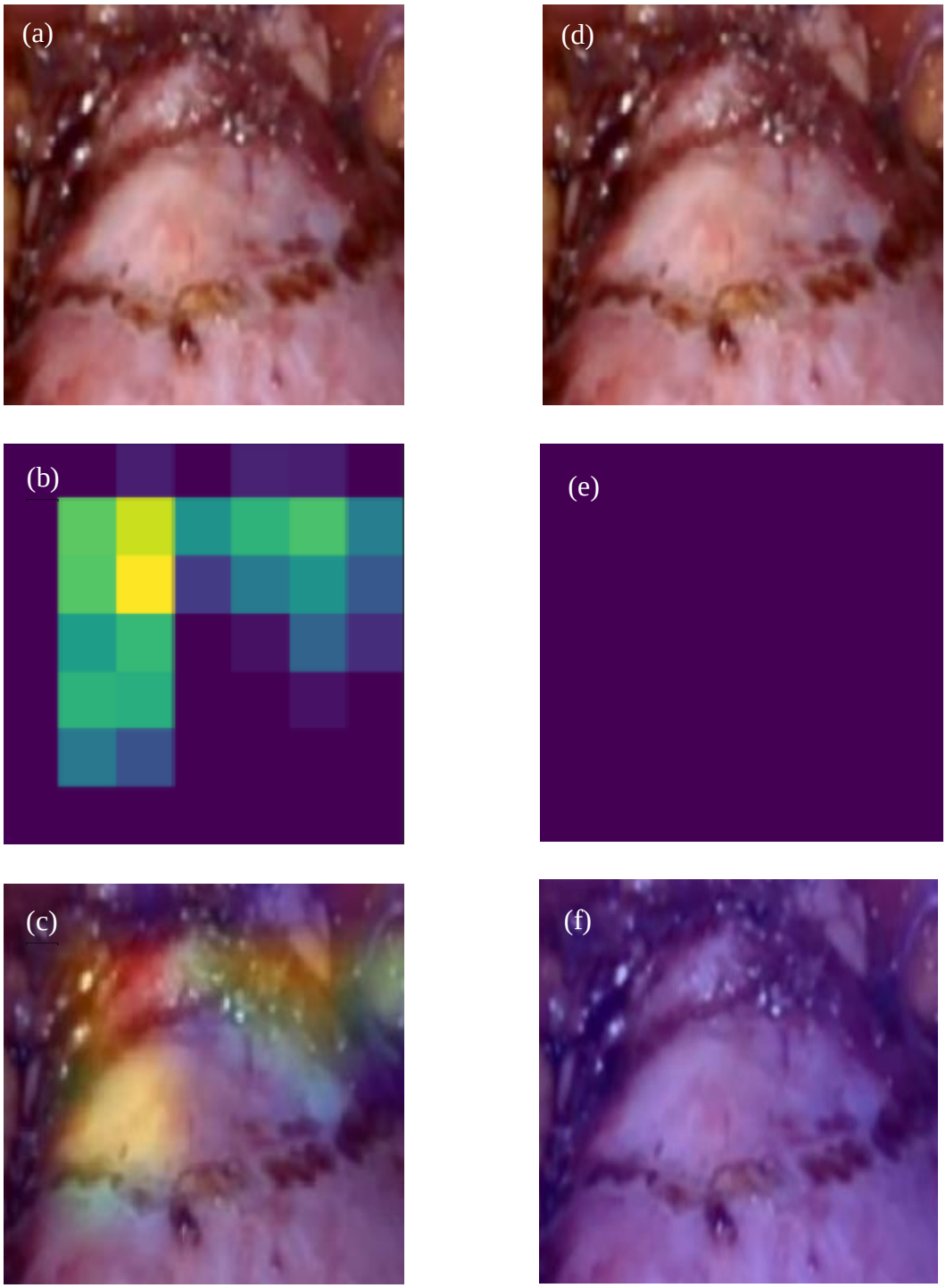


Figure29: Comparison of VGG-16 with lower lr, dropout, callback (left column) and VGG-16 with lower lr, callback (right column) for 5th dataset (3 class classification). (a) One of the images from the test set, (b) Coarse heatmap for the image from lower lr, dropout, callback, (c) Heatmap for the image from lower lr, dropout, callback. On the right column it was not detected. (d) The same image from the test set, (e) Network was not able to detect the image, that is why it is purple (f) No heatmap got detected on the image.

From Figure 29, VGG16 5th dataset lower lr Dropout 0.5 callback was the model that was selected for 3 class classification.

Comparison was done the same way for 2 class classification and VGG-16 6th dataset 2 lower lr using callback was selected as the model for 2 class classification. Look into appendix C for further comparison. Next, Class Activations was used as another evaluation metric to validate that the selected models would be suitable for 2 class and 3 class classification.

4.3.5 Class Activations (Feature maps)

Class activations or feature maps are what each of the neuron's output in a network. If any of the neurons are inactive in a network, they are left as purple (See Appendix D for an example). It is optimal to have as much as active neurons in the network as possible. Inactive neurons tend to overfit the network and the model does not generalize very well. Table 10 (5th dataset) and 11 (6th dataset) shows the number of inactive neurons present in each layer for the models used for 5th dataset and 6th dataset. The network that will have the least number of inactive neurons will be the most suitable one.

Table 10: Number of inactive neurons in each layer of these networks for 5th dataset

Layer name	**VGG-16**	**VGG-16 lower lr**	**VGG-16 lower lr callback**	**VGG-16 lower lr dropout 0.5**	**VGG-16 lower lr dropout 0.5 callback**
conv2d (64)	6	7	7	7	7
conv2d_1 (64)	1	1	3	3	3
max_pooling2d (64)	1	1	3	3	3
conv2d_2 (128)	3	8	11	7	9
conv2d_3 (128)	1	4	3	6	4
max_pooling2d_1 (128)	1	4	3	6	4
conv2d_4 (256)	2	5	6	8	8
conv2d_5 (256)	8	3	8	7	3
conv2d_6 (256)	3	5	5	4	6
max_pooling2d_2 (256)	3	5	5	4	6
conv2d_7 (512)	32	9	9	11	7
conv2d_8 (512)	79	6	6	7	5
conv2d_9 (512)	63	6	7	5	5
max_pooling2d_3 (512)	63	6	7	5	5
conv2d_10 (512)	55	6	15	8	10
conv2d_11 (512)	53	7	5	7	6
conv2d_12 (512)	155	1	1	1	2
max_pooling2d_4 (512)	155	1	1	1	2
Total	684	85	105	100	95

Table 10 shows an indication that for 3 class classification VGG-16 lower lr dropout 0.5 callback or VGG-16 lower lr would be better. But from previous results it was confirmed that the performance

of VGG-16 lower lr is not good for 3 class classification. Therefore, for 3 class classification, VGG-16 lower lr dropout 0.5 callback would be best.

Now, the same method has been used for confirming the best neural network for 2 class classification. Counting the inactive neurons in the networks layer by layer and tabulating the numbers for further comparison. Table 11 below shows the number of inactive neurons for 2 class classification networks.

Table 11: Number of inactive neurons in each layer of these networks for 6th dataset

Layer name	**VGG-16**	**VGG-16 lower lr**	**VGG-16 lower lr callback**	**VGG-16 lower lr dropout 0.5**	**VGG-16 lower lr dropout 0.5 callback**
conv2d	5	6	6	6	6
conv2d_1	1	5	5	6	5
max_pooling2d	1	5	5	6	5
conv2d_2	1	10	10	10	10
conv2d_3	0	6	8	7	7
max_pooling2d_1	0	5	8	7	7
conv2d_4	3	13	14	16	14
conv2d_5	0	17	15	15	17
conv2d_6	4	21	21	23	24
max_pooling2d_2	4	21	21	23	24
conv2d_7	9	51	49	46	48
conv2d_8	55	27	26	31	34
conv2d_9	48	31	26	35	32
max_pooling2d_3	48	31	26	35	32
conv2d_10	27	51	52	53	56
conv2d_11	34	29	31	34	30
conv2d_12	102	10	12	19	12
max_pooling2d_4	102	10	12	19	12
Total	444	349	347	391	375

From Table 11 it is confirmed that the VGG-16 lower lr callback which provided good performance in terms of other evaluation metrics also performed the best in terms of class activations (Feature maps).

The comparison of this study with other related studies are summarized in the Table 11 below-

Table 11: Comparison with other studies

Method	**Image type**	**AI technique used**	**Total images (TI)**	**Evaluation metric**	**Validation performance (VP)**	$\frac{VP}{TI}$
Hadjiyski et al	CT scans	Inception v3	4200	AUC	86%	0.02
Aubreville et al.	Whole Slide Images	RetinaNet with ResNet-50	13,907	F1 score	79.1%	0.01
Wang et al	Multi parametric MRI	V-net	79 cases in total. About 790 images.	Accuracy	89.4%	0.11
Chung et al	Multi parametric MRI	SVM with RD-CRF	20 cases in total. About 200 images.	Accuracy	59%	0.29
Brunese et al	Chest X-ray	VGG-16	9326	Accuracy	98%	0.01
Wu eta al	Chest CT scan	VGG-16 with segmentation	3,855	Sensitivity	95%	0.03
This study	Live partial robotic nephrectomy	Object detection with VGG-16	143	Accuracy	84%	0.59

5. Conclusion and Future Work

Considering there is no live tumor detection technology currently in the da Vinci Xi robots, this project proposed a CNN approach to help surgeons detect tumors real-time during surgery. Global range tumor detection inside the patient was done via YOLOv4. The close range detection approach was built on VGG-16 base model. Two main models were considered for the project. Variation of the models were also considered. For global range detection, there was comparison between YOLOv3 and YOLOv4. For classification, comparison was between 2 classes (Cancerous tissue, Non cancerous tissue) and 3 classes (Cancerous tissue, Fatty tissue, Non cancerous tissue) and for the 2 variations, 5 different models of VGG-16 were considered. The other model classified between 2 classes which included cancerous and non-cancerous tissue. Also, the areas where tumor was detected was highlighted depending on the output of the CNN model (more details of this in the Appendix C).

For 2 class classification, with 150 cancerous tissue images and 150 non-cancerous tissue images in the training set, the final accuracy was 0.84. For 3 class classification, with 105 images for each of cancerous, non-cancerous and fatty tissue in the training set, the final accuracy was 0.69. The proposed method was for identifying tumors in global range at first, and then, when the tumor had been cut off, close range (In methodology section it was mentioned that the images were cropped to include the cancerous and non-cancerous portion) detection would come into play to give the surgeons a second opinion in terms of identifying if there were any more tumors that the surgeon had missed. Looking at the results in this project, it is hoped that surgeons will be more interested in making dataset on real-time surgery images available online. Provided about 5000 images (1500 cancerous tissue for training, 1500 non-cancerous tissue for training, 500 cancerous tissue for validation, 500 non-cancerous tissue for validation, 500 cancerous tissue for testing, 500 non-cancerous tissue for testing) can be made available, it will enable the results to be more promising and will allow detection on a more customized scale.

References

[1] T. h. E. Team. (October 24, 2017). *What do you want to know about cancer*. Available: https://www.healthline.com/health/cancer#growth

[2] C. Council. *What is cancer*. Available: https://www.cancer.org.au/cancer-information/what-is-cancer

[3] A. C. R. Foundation. *Prostate Cancer*. Available: https://www.acrf.com.au/support-cancer-research/types-of-cancer/prostate-cancer/?gclid=Cj0KCQiA7NKBBhDBARIsAHbXCB7I5TPCn_rlyGLLNYbCcRmKkJj3C1jdcim-35do2_NGWPRCfabdt44aAuC7EALw_wcB

[4] P. c. UK. (2018). *Rare prostate cancers*. Available: https://prostatecanceruk.org/prostate-information/further-help/rare-prostate-cancer#:~:text=Like%20common%20prostate%20cancer%2C%20some,may%20hear%20them%20called%20adenocarcinomas

[5] C. council. *Prostate cancer*. Available: https://www.cancer.org.au/cancer-information/types-of-cancer/prostate-cancer

[6] H. Moradi, S. Tang, and S. E. Salcudean, "Toward Intra-Operative Prostate Photoacoustic Imaging: Configuration Evaluation and Implementation Using the da Vinci Research Kit," *IEEE Transactions on Medical Imaging,* vol. 38, no. 1, pp. 57-68, 2019.

[7] S. G. Urology. *Kidney cancers*. Available: https://www.stgeorgeurology.com.au/kidney-cancers

[8] C. C. Australia. (November 2020). *Understanding Kidney Cancer, Last medical review of source booklet*. Available: https://www.cancer.org.au/assets/pdf/understanding-kidney-cancer-booklet

[9] A. Giorgi. *What is Laparoscopy*. Available: https://www.healthline.com/health/laparoscopy

[10] Intuitive. *About da Vinci systems*. Available: https://www.davincisurgery.com/da-vinci-systems/about-da-vinci-systems

[11] J. Park, W. J. Park, C. Lee, M. Kim, S. Kim, and H. J. Kim, "Endoscopic Camera Manipulation planning of a surgical robot using Rapidly-Exploring Random Tree algorithm," in *2015 15th*

International Conference on Control, Automation and Systems (ICCAS), 2015, pp. 1516-1519.
[12] T. Jewell. (September 18, 2018). *What is Minimally Invasive Surgery*. Available: https://www.healthline.com/health/minimally-invasive-surgery#conditions-treatedwith-robotic-surgery
[13] A. S. Walker and S. R. Steele, "The future of robotic instruments in colon and rectal surgery," *Seminars in Colon and Rectal Surgery*, vol. 27, no. 3, pp. 144-149, 2016/09/01/ 2016.
[14] Intuitive. *da Vinci vision*. Available: https://www.intuitive.com/en-us/products-and-services/da-vinci/vision
[15] Intuitive. *da Vinci Instruments*. Available: https://www.intuitive.com/en-us/products-and-services/da-vinci/instruments
[16] J. Lin, E. L. Knight, M. L. Hogan, and A. K. Singh, "A Comparison of Prediction Equations for Estimating Glomerular Filtration Rate in Adults without Kidney Disease," *Journal of the American Society of Nephrology*, vol. 14, no. 10, p. 2573, 2003.
[17] A. Geron, *Hands-on Machine Learning with Scikit-Learn, Keras & TensorFlow*. o'Reiley Media, Inc, September, 2019.
[18] D. H. Hubel, "Single Unit Activity in Striate Cortex of Unrestrained Cats," *The Journal of Physiology*, vol. 147, pp. 226-238, September 2, 1959.
[19] D. H. Hubel and T. N. Wiesel, "Receptive fields of single neurones in the cat's striate cortex," (in eng), *J Physiol*, vol. 148, no. 3, pp. 574-91, Oct 1959.
[20] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biological Cybernetics*, vol. 36, no. 4, pp. 193-202, 1980/04/01 1980.
[21] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998.
[22] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," vol. 60, no. 6 %J Commun. ACM, pp. 84–90, 2017.
[23] Imagenet. *Dataset*. Available: https://image-net.org/
[24] M. D. Zeiler and R. Fergus, "Visualizing and Understanding Convolutional Networks," in *Computer Vision – ECCV 2014*, Cham, 2014, pp. 818-833: Springer International Publishing.
[25] K. Simonyan and A. Zisserman, " Very Deep Convolutional Networks for Large-Scale Image Recognition," 4 September, 2014.
[26] R. C. D. Capper. (25 March, 2020). *How AI is improving cancer diagnosis*. Available: https://www.nature.com/articles/d41586-020-00847-2
[27] A. G. Chung, F. Khalvati, M. J. Shafiee, M. A. Haider, and A. Wong, "Prostate Cancer Detection via a Quantitative Radiomics-Driven Conditional Random Field Framework," *IEEE Access*, vol. 3, pp. 2531-2541, 2015.
[28] L. Pantanowitz *et al.*, "An artificial intelligence algorithm for prostate cancer diagnosis in whole slide images of core needle biopsies: a blinded clinical validation and deployment study," *The Lancet Digital Health*, vol. 2, no. 8, pp. e407-e416, 2020.
[29] Y. Wang, B. Zheng, D. Gao, and J. Wang, "Fully convolutional neural networks for prostate cancer detection using multi-parametric magnetic resonance images: an initial investigation," in *2018 24th International Conference on Pattern Recognition (ICPR)*, 2018, pp. 3814-3819.
[30] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision (3DV)*, 2016, pp. 565-571: IEEE.
[31] G. H. Aly, M. Marey, S. A. El-Sayed, and M. F. Tolba, "YOLO Based Breast Masses Detection and Classification in Full-Field Digital Mammograms," *Computer Methods and Programs in Biomedicine*, vol. 200, p. 105823, 2021/03/01/ 2021.
[32] H. M. Ünver and E. Ayan, "Skin Lesion Segmentation in Dermoscopic Images with Combination of YOLO and GrabCut Algorithm," (in eng), *Diagnostics (Basel)*, vol. 9, no. 3, Jul 10 2019.

[33] N. Tangri *et al.*, "Risk prediction models for patients with chronic kidney disease: a systematic review," (in eng), *Ann Intern Med*, vol. 158, no. 8, pp. 596-603, Apr 16 2013.
[34] M. Aubreville, C. A. Bertram, T. A. Donovan, C. Marzahl, A. Maier, and R. Klopfleisch, "A completely annotated whole slide image dataset of canine breast cancer to aid human breast cancer research," *Scientific Data*, vol. 7, no. 1, p. 417, 2020/11/27 2020.
[35] L. Brunese, F. Mercaldo, A. Reginelli, A. J. C. M. Santone, and P. i. Biomedicine, "Explainable Deep Learning for Pulmonary Disease and Coronavirus COVID-19 Detection from X-rays," vol. 196, pp. 105608 - 105608, 2020.
[36] H. V. Huff and A. Singh, "Asymptomatic Transmission During the Coronavirus Disease 2019 Pandemic and Implications for Public Health Strategies," (in eng), *Clin Infect Dis*, vol. 71, no. 10, pp. 2752-2756, Dec 17 2020.
[37] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 618-626.
[38] M. Li, Z. Zhang, L. Lei, X. Wang, and X. Guo, "Agricultural Greenhouses Detection in High-Resolution Satellite Images Based on Convolutional Neural Networks: Comparison of Faster R-CNN, YOLO v3 and SSD," *Sensors*, vol. 20, no. 17, 2020.
[39] L. Ge, D. Dan, and L. Hui, "An accurate and robust monitoring method of full-bridge traffic load distribution based on YOLO-v3 machine vision," *Structural Control and Heath Monitoring*, vol. 27, 24 September, 2020.
[40] Z. Chen, T. Zhang, and C. Ouyang, "End-to-End Airplane Detection Using Transfer Learning in Remote Sensing Images," *Remote Sensing*, vol. 10, no. 1, 2018.
[41] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 3431-3440.
[42] V. Granata *et al.*, "Immediate Adverse Reactions to Gadolinium-Based MR Contrast Media: A Retrospective Analysis on 10,608 Examinations," (in eng), *Biomed Res Int*, vol. 2016, p. 3918292, 2016.
[43] INbreast. *Dataset*. Available: http://medicalresearch.inescporto.pt/breastresearch/index.php/Get_INbreast_Database
[44] T. Lee, V. Ng, R. Gallagher, A. Coldman, and D. McLean, "DullRazor: a software approach to hair removal from images," (in eng), *Comput Biol Med*, vol. 27, no. 6, pp. 533-43, Nov 1997.
[45] C. Rother, V. Kolmogorov, and A. Blake, ""GrabCut": interactive foreground extraction using iterated graph cuts," vol. 23, no. 3 %J ACM Trans. Graph., pp. 309–314, 2004.
[46] tzulatin. (2017). *LabelImg[source code]*. Available: https://github.com/tzutalin/labelImg#labelimg
[47] A. Bochkovskiy, C. Wang, and H. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," *Computer Vison and Pattern Recognition*, vol. 1, April 23, 2020.
[48] S. S. Foundation. (Febrary 29, 2020). *Robotic Partial Nephrectomy for Complex Tumors presented by Ronney Abaza*. Available: https://www.youtube.com/watch?v=vvf16vBrgxQ&t=662s&ab_channel=SeattleScienceFoundation
[49] B. A. W. s. Hospital. (December 12, 2018). *Robotic Assisted Laparoscopic Partial Nephrectomy*. Available: https://www.youtube.com/watch?v=GQm90mWVMJM&ab_channel=BrighamAndWomen%27sHospital
[50] S. S. Foundation, "LIVE SURGERY: Retroperitoneal Robotic Partial Nephrectomy," September 30, 2015. Available: https://www.youtube.com/watch?v=nwrbKNbLCv8&t=5045s&ab_channel=SeattleScienceFoundation

[51] P. N. U. Specialist, "Robotic partial nephrectomy comparisons," January 25, 2020. Available: https://www.youtube.com/watch?v=epvKkH3ekRo&ab_channel=PacificNorthwestUrologySpecialists%2CPLLC

[52] V. Foundation, "Dr. Craig Rogers: da Vinci Partial Nephrectomy," July 7,2015. Available: https://www.youtube.com/watch?v=gdg7EhsKki8&ab_channel=VattikutiFoundation

[53] J. D. E. M.D., "Robotic Partial Nephrectomy," March 16, 2016.Available: https://www.youtube.com/watch?v=UXWjNqTwb_4&ab_channel=JasonD.Engel%2CM.D.

[54] GlobalCastMD. (October 11, 2014). *02 Robotic Partial NephrectomyCourse Tips for retroperitoneral partial nephrectomy James Porter HD*. Available: https://www.youtube.com/watch?v=S80t7cnFLus&ab_channel=GlobalCastMD

[55] S. S. Foundation. (February 29, 2020). *Avoiding Posotive Margins During Robotic Partial Nephrectomy presented by Ronney Abaza*. Available: https://www.youtube.com/watch?v=C3VTbb_1GAM&ab_channel=SeattleScienceFoundation

[56] VCUrobotics. (October 21, 2015). *da Vinci Xi Right Robotic Partial Nephrectomy-Unedited*. Available: https://www.youtube.com/watch?v=6eyZzoScc54&ab_channel=VCUrobotics

[57] AlexeyAB. *darknet[source code]*. Available: https://github.com/AlexeyAB/darknet

[58] A. NG. *Machine Learning, Bias VS Variance*. Available: https://www.coursera.org/learn/machine-learning

[59] B. David, M. D. Samadi. *History and The Future of Robotic Surgery*, Robotic Oncology. [online]. Available: https://www.roboticoncology.com/history-of-robotic-surgery/

[60] American Institute of Minimally Invasive Surgery. *DA VINCI XI*, American Medical Center,2019. Accessed on: June 26, 2021. [online]. Available: https://www.aimisrobotics.com/da-vinci-xi/

[61] A. Inc, B. Grove, *Indocyanine green Side Effects*, March 19, 2021. Available: https://www.drugs.com/sfx/indocyanine-green-side-effects.html

[62] A. G. Chung, F. Khalvati, M. J. Shafiee, M. A. Haider, and A. Wong, "Prostate Cancer Detection via a Quantitative Radiomics-Driven Conditional Random Field Framework," *IEEE Access*, vol. 3, pp. 2531-2541, 2015.

[63] M. D. Zeiler and R. Fergus, "Visualizing and Understanding Convolutional Networks," in *Computer Vision – ECCV 2014*, Cham, 2014, pp. 818-833: Springer International Publishing.

[64] H. Shin *et al.*, "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285-1298, 2016.

[65] L. Pantanowitz *et al.*, "An artificial intelligence algorithm for prostate cancer diagnosis in whole slide images of core needle biopsies: a blinded clinical validation and deployment study," *The Lancet Digital Health*, vol. 2, no. 8, pp. e407-e416, 2020.

[66] M. Aubreville, C. A. Bertram, T. A. Donovan, C. Marzahl, A. Maier, and R. Klopfleisch, "A completely annotated whole slide image dataset of canine breast cancer to aid human breast cancer research," *Scientific Data*, vol. 7, no. 1, p. 417, 2020/11/27 2020.

[67] M. Li, Z. Zhang, L. Lei, X. Wang, and X. Guo, "Agricultural Greenhouses Detection in High-Resolution Satellite Images Based on Convolutional Neural Networks: Comparison of Faster R-CNN, YOLO v3 and SSD," *Sensors*, vol. 20, no. 17, 2020.

[68] L. Ge, D. Dan, and L. Hui, "An accurate and robust monitoring method of full-bridge traffic load distribution based on YOLO-v3 machine vision," *Structural Control and Heath Monitoring*, vol. 27, 24 September, 2020.

[69] Z. Chen, T. Zhang, and C. Ouyang, "End-to-End Airplane Detection Using Transfer Learning in Remote Sensing Images," *Remote Sensing*, vol. 10, no. 1, 2018.

[70] G. Kharate, A. J. I. J. o. S. S. Ghotkar, and I. Systems, "VISION BASED MULTI-FEATURE HAND GESTURE RECOGNITION FOR INDIAN SIGN LANGUAGE MANUAL SIGNS," International Journal on Smart Sensing and Intelligent Systems, vol. 9, pp. 124-147, 2016.

[71] D. Nakhaeinia *et al.*, "SURFACE FOLLOWING WITH AN RGB-D VISION-GUIDED ROBOTIC SYSTEM FOR AUTOMATED AND RAPID VEHICLE INSPECTION," International Journal on Smart Sensing and Intelligent Systems, vol. 9, pp. 419-447, 2016.

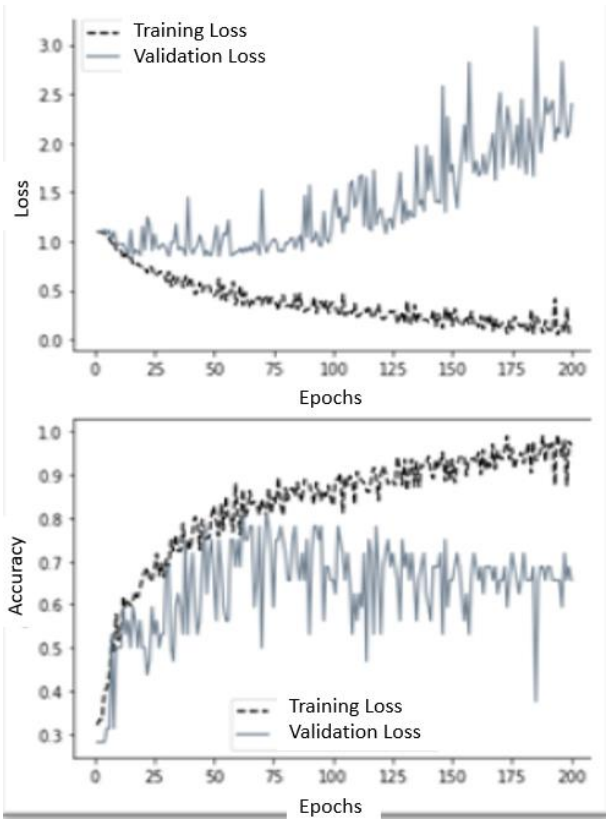
[72] M. Bennet, B. Thamilvalluvan, P. P. Alphonse, D. R. Thendralarasi, K. J. I. J. o. S. S. Sujithra, and I. Systems, "PERFORMANCE AND ANALYSIS OF AUTOMATIC LICENSE PLATE LOCALIZATION AND RECOGNITION FROM VIDEO SEQUENCES," International Journal on Smart Sensing and Intelligent Systems, vol. 10, pp. 330-343, 2017.

Appendix A

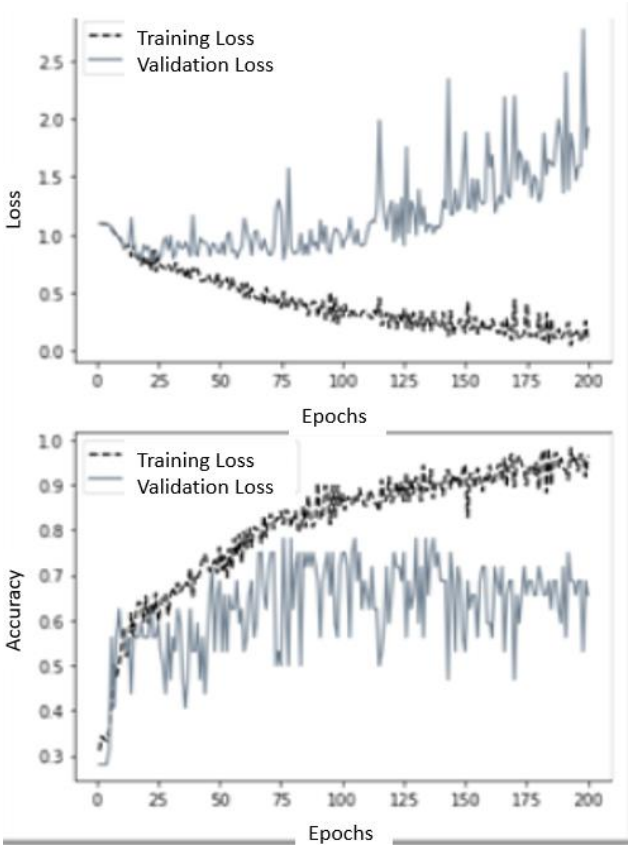
Performance evaluation

- Loss VS epochs
- Confusion matrix
- Accuracy
- GRAD-CAM
- Class activations

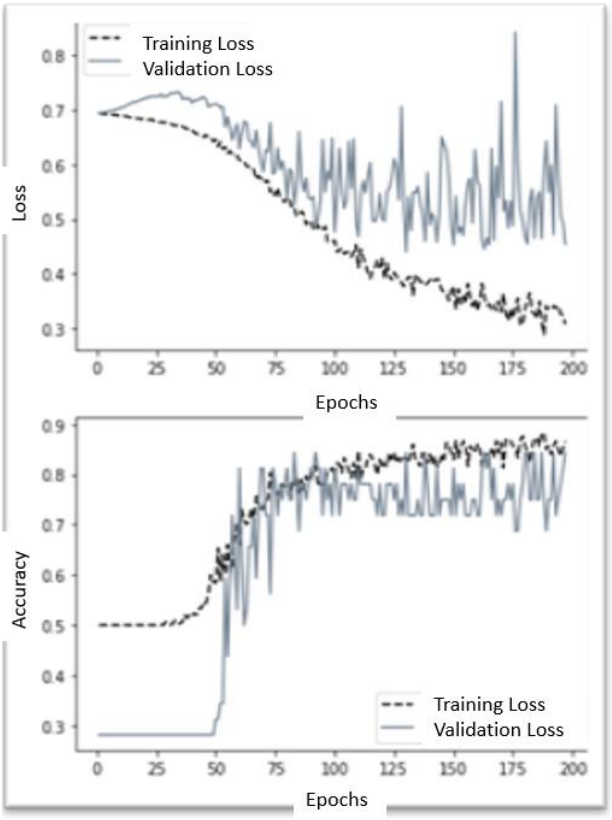
Loss VS epochs



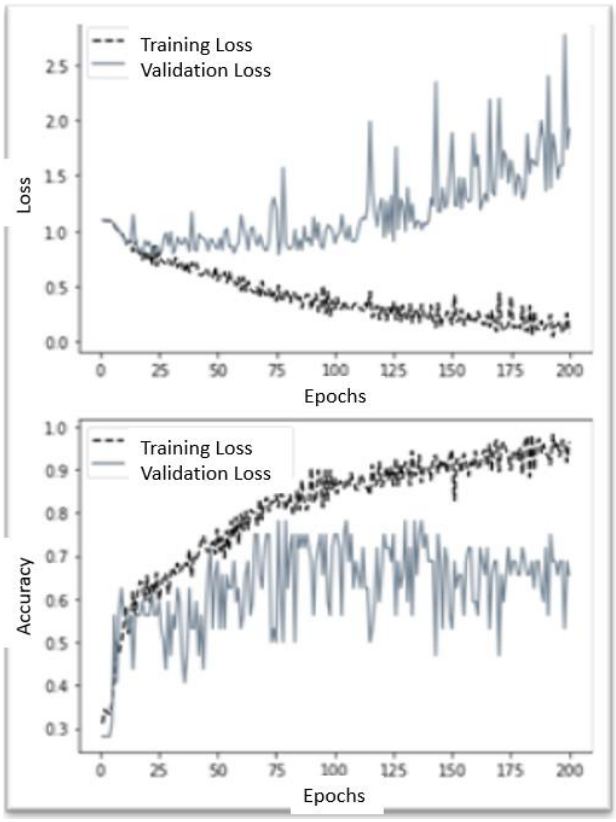
A1: VGG-16 5th dataset lower lr callback



A2: VGG-16 5th dataset lower lr dropout 0.5

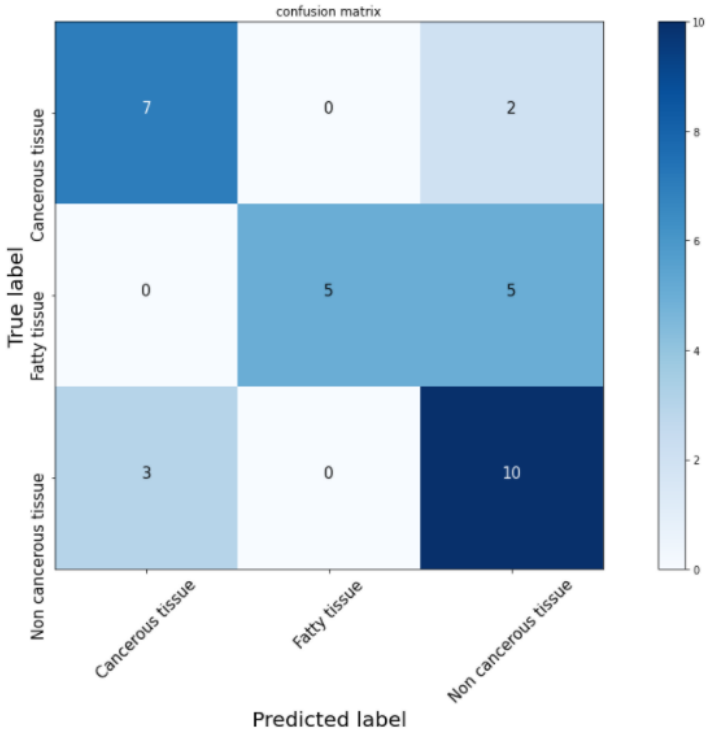


A3: VGG-16 6th dataset lower lr callback

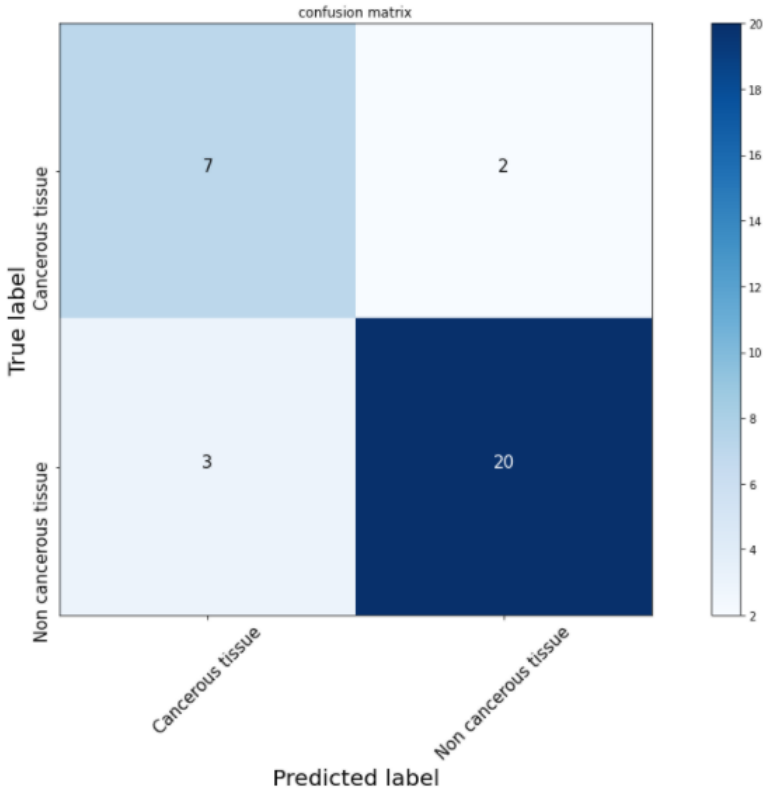


A4: VGG-16 6th dataset lower lr

Appendix B



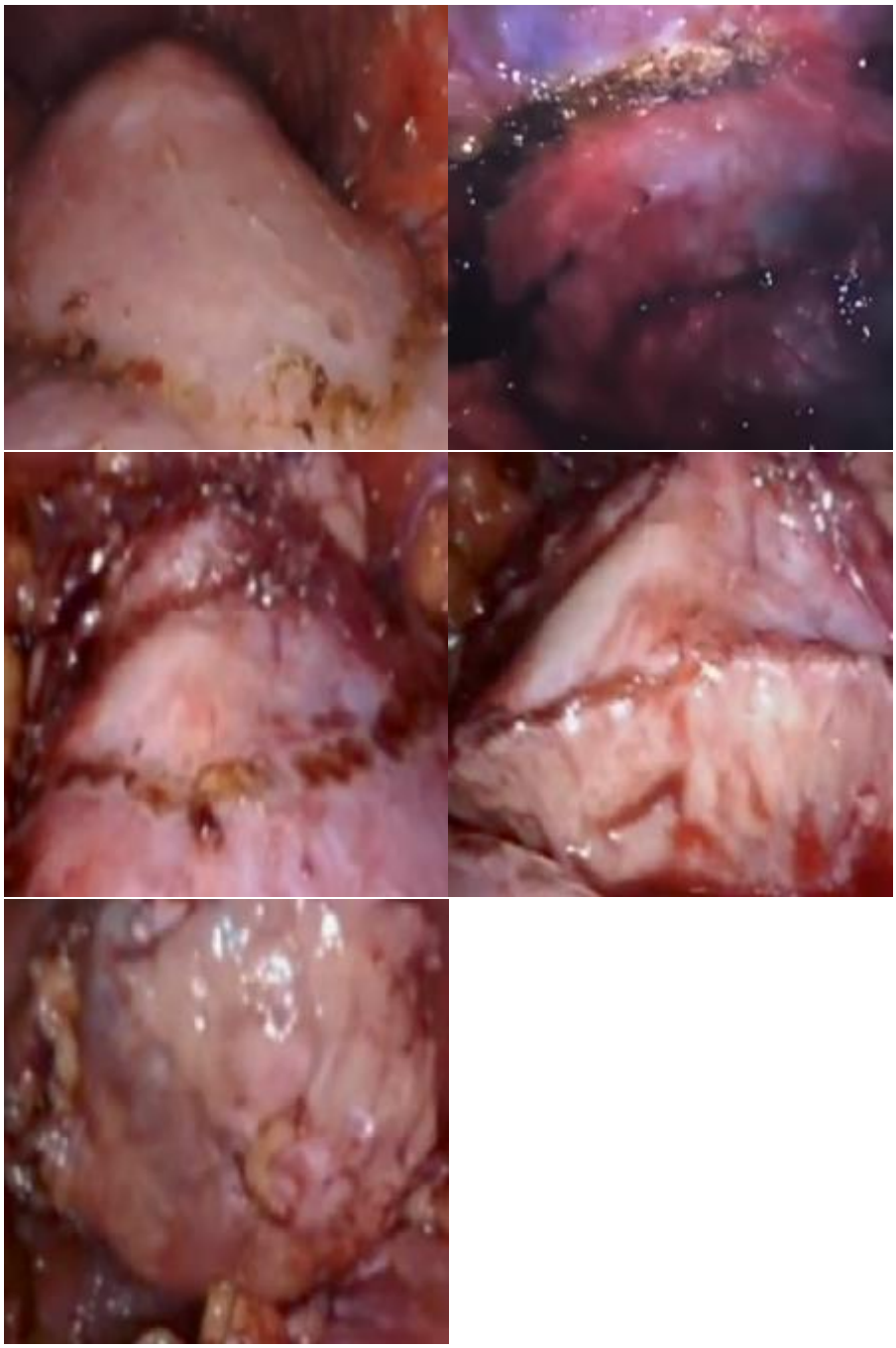
B1: VGG-16 5th dataset lower lr callback



B2: VGG-16 6th dataset lower lr dropout 0.5

Appendix C

There were 5 cancerous tissue images in the test set. They are as below-

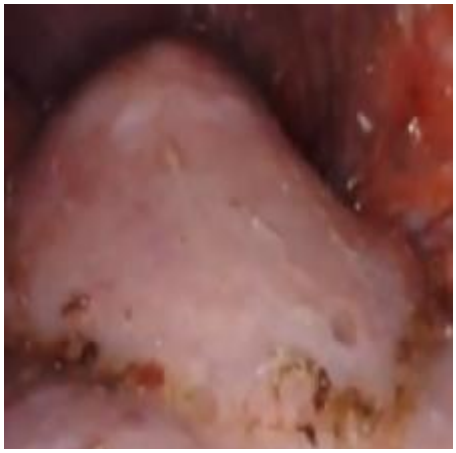


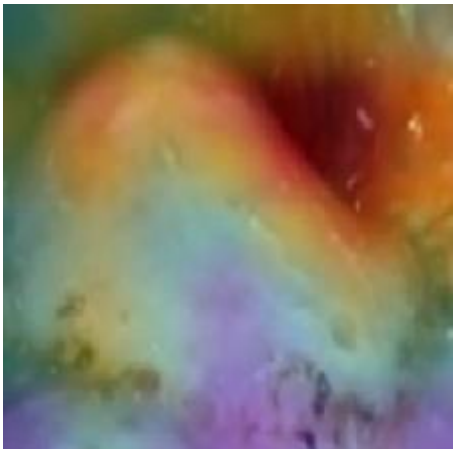
After the images had been tested with the model where 3 classes were considered, the model out outputs 3 probabilities. The output for each of the images are presented below along with their heatmaps highlighting which portion of the image was considered to give the prediction.

The model outputs 3 probabilities. The first number is for cancerous tissue, the second is for fatty tissue, the third is for non-cancerous tissue.

Visualizing heatmaps and prediction outputs

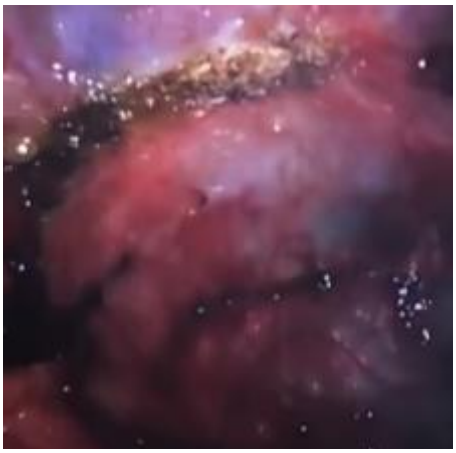
5th dataset output





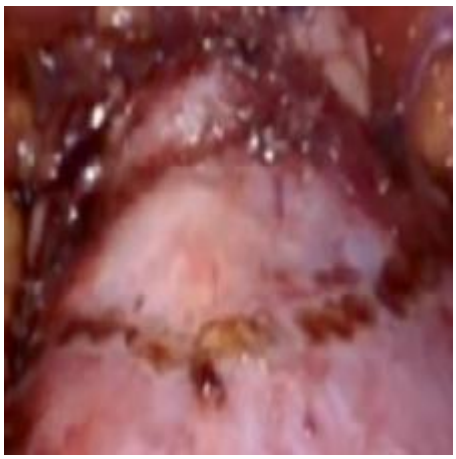
(0.018, 0.00009,0.982)

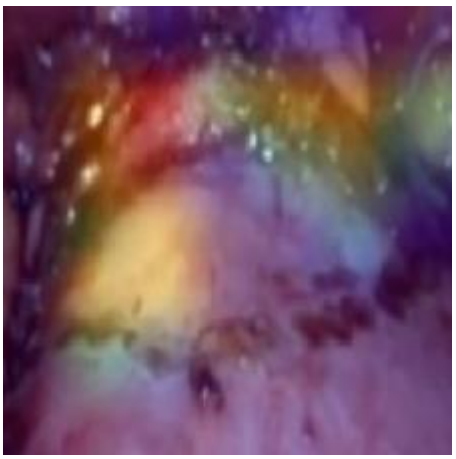
More probability for non-cancerous tissue. The red highlights the non-cancerous portion. The blue and the purple is for cancerous portion.



(0.840, 0.00019,0.159)

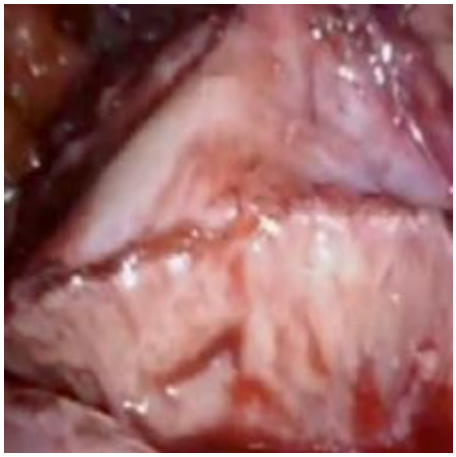
More probability for cancerous tissue. The red highlights the cancerous portion. The blue and the purple is for non-cancerous portion.

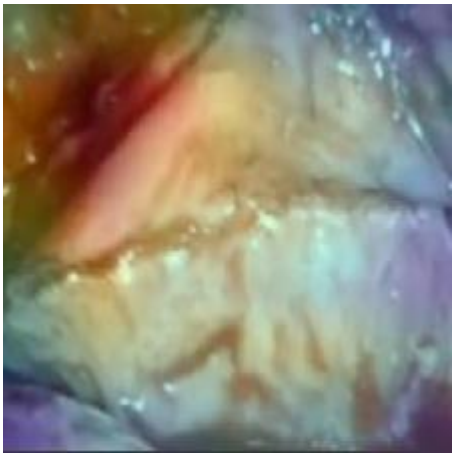




(0.971,0.00035,0.028)

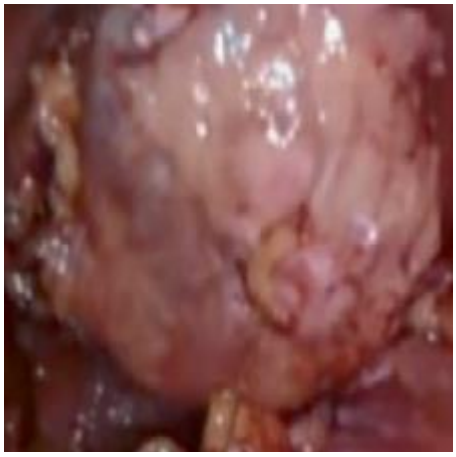
More probability for cancerous tissue. The red highlights the cancerous portion. The blue and the purple is for non-cancerous portion.

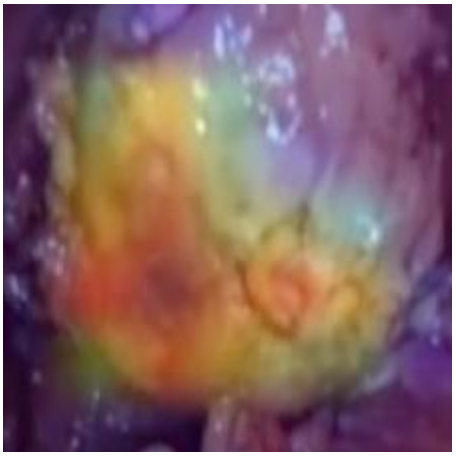




(0.953,0.00018,0.047)

More probability for cancerous tissue. The red highlights the cancerous portion. The blue and the purple is for non-cancerous portion.





(0.286,0.00059,0.714)

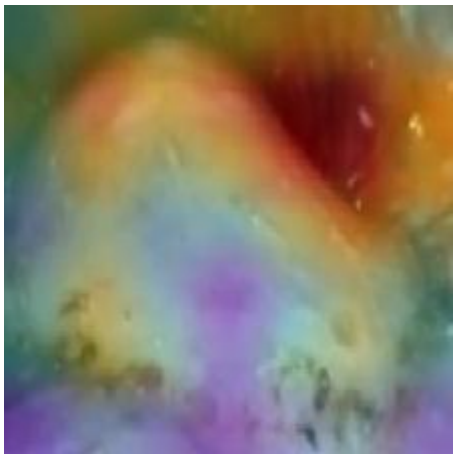
More probability for non-cancerous tissue. The red highlights the non-cancerous

portion. The blue and the purple is for cancerous portion.

6th dataset output

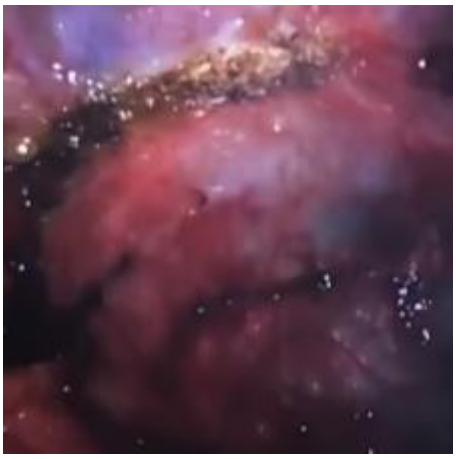
After the images had been tested with the model where 2 classes were considered, the model outputs 2 probabilities. The output for each of the images are presented below along with their heatmaps highlighting which portion of the image was considered to give the prediction.

The model outputs 2 probabilities. The first number is for cancerous tissue, the second is for non-cancerous tissue.



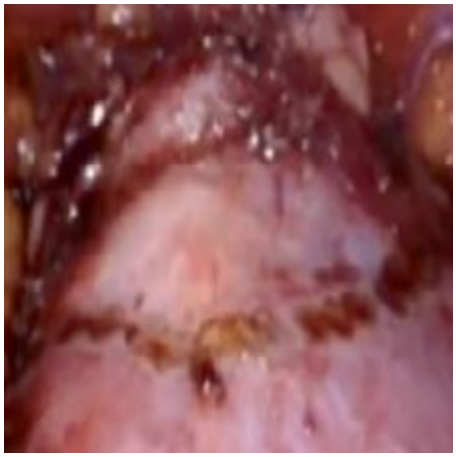
(0.131,0.869)

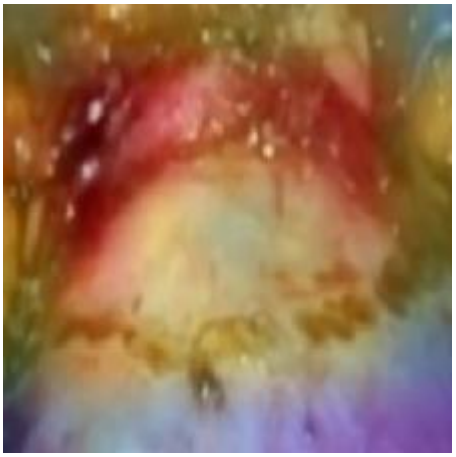
More probability for non-cancerous tissue. The red highlights the non-cancerous portion. The blue and the purple is for cancerous portion.



(0.046,0.954)

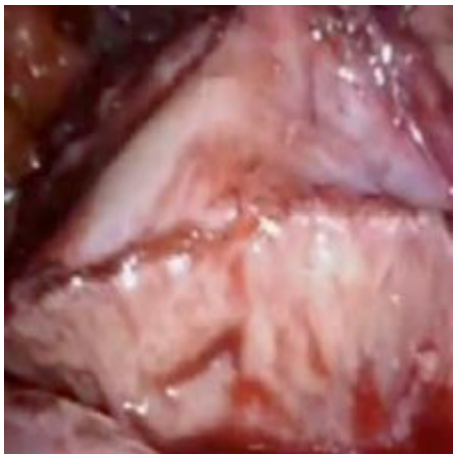
More probability for non-cancerous tissue. The red highlights the non-cancerous portion. The blue and the purple is for cancerous portion. This one is wrong.

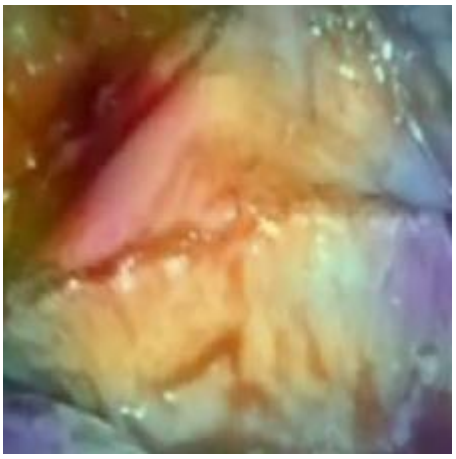




(0.667,0.333)

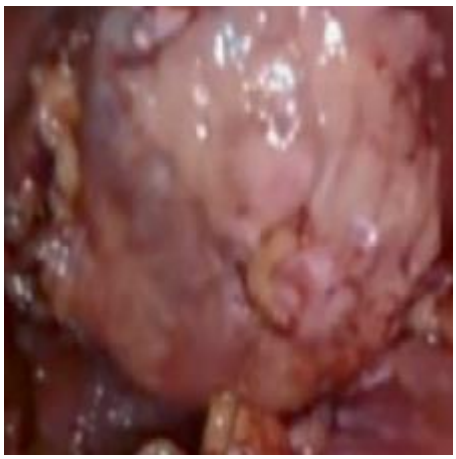
More probability for cancerous tissue. The red highlights the cancerous portion. The blue and the purple is for non-cancerous portion.

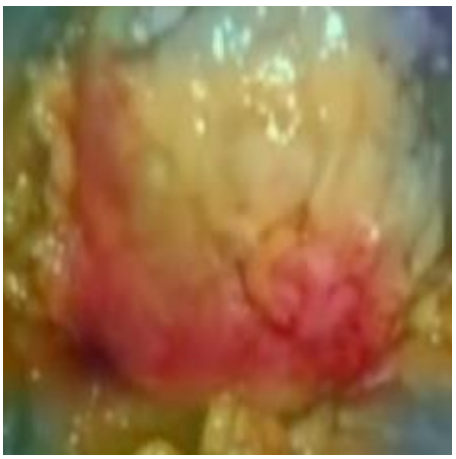




(0.880,0.120)

More probability for cancerous tissue. The red highlights the cancerous portion. The blue and the purple is for non-cancerous portion.

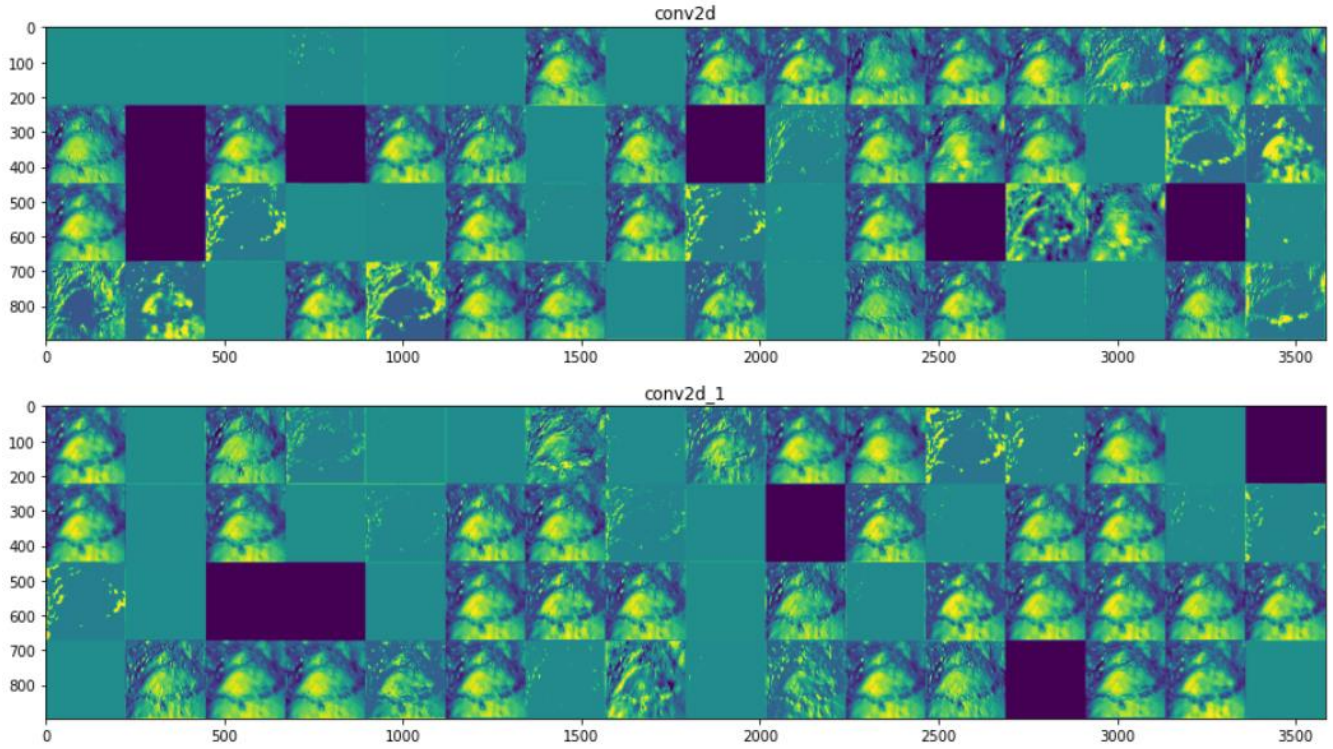




(0.280,0.720)

More probability for non-cancerous tissue. The red highlights the non-cancerous portion. The blue and the purple is for cancerous portion.

Appendix D



D1: Class activation for the 1st 2 layers of VGG-16. The purple regions are inactive neurons