MASTER

Tissue recognition for contrast enhanced ultrasound videos

Singhi, S.

Award date:
2018

Awarding institution:
Technische Universität Berlin

Disclaimer
This document contains a student thesis (bachelor's or master's), as authored by a student at Eindhoven University of Technology. Student theses are made available in the TU/e repository upon obtaining the required degree. The grade received is not published on the document as presented in the repository. The required complexity or quality of research of student theses may vary by program, and the required minimum study period may vary in duration.

Tissue Recognition
for Contrast Enhanced Ultrasound Videos

Tissue Recognition
for Contrast Enhanced Ultrasound Videos

Samridhi Singhi

A thesis submitted to the
Faculty of Electrical Engineering and Computer Science
of the
Technical University of Berlin
in partial fulfillment of the requirements for the degree
Master of Data Science

Berlin, Germany
September 30, 2018

Main supervisor:

Prof. Dr. habil. Odej Kao, Technical University of Berlin

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und eigenhändig sowie ohne unerlaubte fremde Hilfe und ausschließlich unter Verwendung der aufgeführten Quellen und Hilfsmittel angefertigt habe.

Berlin, den

Abstract

Sonograph imaging or Ultrasound is a rapid, non-invasive method used as a diagnostic step to identify surgical emergencies. It is more convenient and less expensive than Magnetic Resonance Imaging or other invasive methods. To further improve diagnostic capabilities, chemicals called Ultrasound Contrast Agents are injected into patients before the Ultrasound process. These chemicals enhance the echogenicity of the blood flow thus improving tissue differentiability. However, Ultrasound requires physicians to undergo significant training and practice to analyse the reports and identify tissues to provide an accurate diagnosis. Artificial Intelligence can aid the physicians to efficiently perform this step by classifying tissues based on their type and health. This thesis explores applying supervised Artificial Intelligence to identify tissues. We analyse contrast enhanced ultrasound videos of 5 patients labelled with information from experts at the University Hospital Münster (UKM). Broadly, the thesis also presents a comparative study on the performance of classical machine learning and deep learning for image classification and object detection. The classifiers built are evaluated with performance metrics like the AUC and intersection over union (IOU) among others. The results show that the Sequential Minimal Optimization (SMO) classifier performs best with an AUC of 0.8 followed by k-NN with an AUC of 0.77. The deep convolutional network - U-Net, built demonstrates a high IOU of 95.89%. Although the U-Net requires longer training duration and heavily depends on underlying hardware, it is found to be more robust and reliable when trained for multiple patients.

Contents

1 Introduction

In 1880 the Curie brothers, became the first to produce ultrasound waves [86]. Medical use of ultrasonic imaging started in the early 20th century after Paul Langevin, a student of the inventors, used it for submarine detection during World War I [61]. Ultrasound or Sonography is one of the key medical imaging techniques omnipresent today for real-time examination of internal organs. Further advancements, backed up with half a century of research resulted in high-intensity focused ultrasound (HIFU), with an intensity of usually 1–3 MHz ultrasound waves, to now enable non-invasive selective tissue necrosis of lesions [33].

Hepatocellular carcinoma (HCC) is the fifth most common cancer worldwide and has caused 782,000 deaths all over the world as recorded by the World Health Organization's International Agency for Research on Cancer in 2012. HCC, the third highest cause of cancer-related deaths, accounts for 7% of all cancers recorded [34]. Liver cirrhosis represents a major risk factor for the development of HCC. The development of a neoplasm in cirrhosis occurs in multiple phases over a long time. This translates to the possibility of detecting different types of nodules in a cirrhotic liver, ranging from regenerative nodules to low-grade dysplastic nodules and high-grade dysplastic nodules. Early diagnosis of HCC enables aggressive treatment, prolonging patients' lives [22].

Techniques like Ultrasound Elastography [36] and Contrast Enhanced Ultrasound further aid the diagnosis of Liver cirrhosis. Elastography eliminates the acoustic similarity of tissues by introducing forces under which the tissues become differentiable. One such force is vibration at a given frequency [71]. In Contrast Enhanced Ultrasound, contrast agents (CA) are injected into the blood streams of the patients prior to the scan. These agents enhance the echogenity of blood flow dynamically depending on the lesion. The normal and affected tissues are then differentiated based on the enhanced blood flow patterns. However, diagnosis of HCC remains a challenge because lesions may be detected as areas of increased enhancement only in the arterial phase, but the short duration of this phase can make full surveillance of the largest organ of the body problematic [20].

Medical imaging examinations are heavily used for diagnosis today. However, a great fraction of these examinations produce normal results, and the detection of only a small number of suspicious lesions by radiologists is considered both difficult and time-consuming. Therefore, evolving from what began as picture archiving and communication systems (PACS), automated computer-aided diagnosis (CAD) has thus become a major research subject in medical imaging and diagnostic radiology. Advances in the field of Computer Science and Artificial Intelligence have further aided in the development of innovative CAD systems [72]. Today, radiologists use CAD results as a ”second opinion”. The purpose of CAD systems is to complement and assist physicians in making diagnosis.

In this thesis, above mentioned challenges are taken into consideration and a CAD scheme for contrast enhanced ultrasound (CEUS) scans of the liver with the aim of lesion tissue recognition is presented. The design of the proposed system is illustrated in Figure 1.1. The CEUS scan videos from the University Hospital Münster (UKM) are collected and fed into the CAD scheme designed. Foremost, the videos are converted to image frames for further analysis. The data collected is also labelled as per information received from the physicians at the hospital. Following steps in the CAD pipeline include pre-processing, feature extraction, classification and finally visualization of tissues identified. The predicted output of the system will assist the radiologists/physicians in diagnosis.

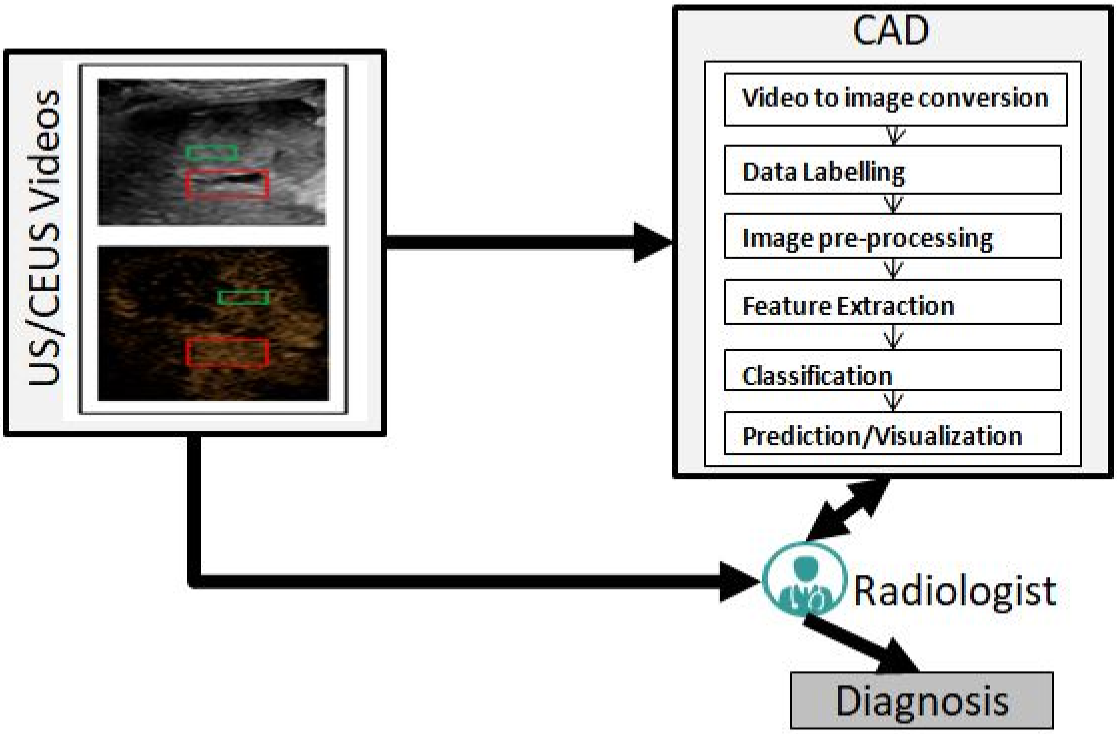


Figure 1.1: Proposed System

Another important contribution of the thesis is the comparison and evaluation of deep-learning and other conventional machine learning techniques involving feature extraction. Both learning techniques have been applied to the dataset collected and the results obtained are evaluated with various evaluation metrics.

The thesis is organized as follows. Chapter 2 explains medical imaging techniques - ultrasound and contrast enhanced ultrasound and industry standards for the same. It sheds light on the medical significance of diagnosis of focal liver lesions. A background on artificial intelligence and evaluation of machine learning algorithms is also presented. In chapter 3 and 4, the main contribution of the thesis are described in detail. Chapter 3 focuses on exploration and analysis of collected data. In chapter 4 we explain the different artificial intelligence approaches taken to implement tissue recognition. The results of the approaches are discussed in chapter 5. The models built are evaluated and the tissue predictions are visualized. Following which, a research on the state of the art on the topic of CAD systems, image processing and object detection is conducted and described in chapter 6. Finally, chapter 7 summarises the findings and concludes the thesis.

2
Background

2.1 Ultrasound

Ultrasound or Sonography is a medical imaging technique. It works on the principal of reflection of sound. This principal can also be observed in nature as seen in the echolocation used by dolphins, whales and bats. One of the first applications of this property of sound led to the invention of sound navigation and ranging (SONAR) systems for submarines during World War I. This later encouraged further research on the applicability of the new technique for medical use.

Interestingly, the use of ultrasonics in the field of medicine began with its use in therapy rather than diagnosis. The disruptive capabilities of high intensity ultrasound had been noticed as early as 1920s by Paul Langévin who also invented the SONAR. High intensity ultrasound progressively evolved to become a neuro-surgical tool. It was also being used in physical and rehabilitation medicine. The 1940s saw ultrasound being claimed as the ”cure all” remedy. However, lacking scientific evidence and increasing concerns about the detection of tissues for treatment and also the harmful effects of ultrasound on neighboring tissues, research quickly shifted to using ultrasound for diagnosis [91].

Karl Theo Dussik, a psychiatrist at the University of Vienna, Austria, is generally regarded as the first physician to have employed ultrasound in medical diagnosis. He attempted to locate brain tumors and the cerebral ventricles by measuring the transmission of ultrasound beam through the skull [12]. Today, the use of ultrasound for diagnosis is omnipresent. The procedure requires no preparation by the patient, is non-invasive, painless and produces real time images.

Figure 2.1 shows the parts of a modern ultrasound machine. The main part of the machine is the transducer probe that sends and receives sound waves. High frequency sound waves of 1-5 megahertz are used for the probing. The probe has one or more piezoelectric quartz crystals which rapidly change their shape when electricity is applied to them. This rapid shape change produces vibrations in the form of sound waves that travel outwards. Similarly, when reflected sound waves hit the crystals, they are converted back to electricity. This phenomena is called the piezoelectric effect. The frequency of the emitted sound waves determine the quality of imaging. This can be controlled by the transducer pulse controls. Although most probing is done by placing these transducers on the body, they can also be inserted into the body for even better imaging.

The Central Processing Unit (CPU) is the brain of the machine. It performs the tasks of sending electrical signals to the transducer and converting electrical signals received back from the transducer to images and storing the images on disk. Based on the time and strength of echoes received, a microprocessor within the CPU assigns each one a position and color, ultimately forming the image known as B-Mode (brightness mode). Ultrasound systems today are commonly described as “real-time” because they have the ability to rapidly display B-Mode images so that any motion that occurs is visualized as and when it happens. Modern systems can show between 15 and 50 images per second. In order to produce the effect of continuous movement, at least 16 frames need to be displayed per second [16]. The stored images are then converted to optical signals and displayed on the monitor. They can be grayscale or color images. Digital Imaging and communication in medicine (DICOM) is a standard that describes how medical image data should be stored, exchanged and printed. This is explained in section 2.4. Controls like the keyboard and cursor can be used to add details to the displayed image on the monitor and they can be printed if needed.

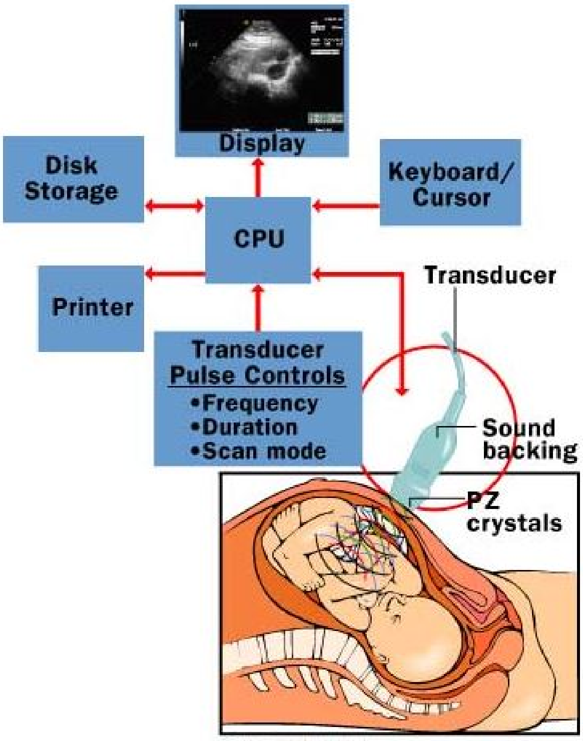


Figure 2.1: Parts of an Ultrasound Machine [17]

The transducer collects the sound waves that are reflected back and converts them to electrical waves. But, not all sound waves sent are reflected back. Some get refracted and scattered [16]. This depends on the angle at which the sound wave is sent but, more importantly, also the surface the wave hits. This helps radiologists identify the tissues on the final image. For instance, Rayleigh Scattering occurs when the wave hits a very small tissue like the red blood cells or arteries or some veins. When a sound wave comes in contact with such small tissues, it scatters creating a uniform amplitude in all directions and hardly reflecting back. Hence, the intensity of such tissues is low on the scan image.

Another possibility is that the sound waves are refracted. When the waves hit the border between two tissues at an oblique angle, due to the difference in density of the media the sound wave travels through, it bends. Useful conclusions about the organs/tissues can also be made by moving the cursor around and observing the images created at each position.

Further, reflection of the sound waves can be of two types- specular or diffuse. Specular reflection occurs when sound hits a large even surface such as bones. Majority of sound waves are sent back and hence the image is most bright. On the other hand, when the sound hits a large soft tissue with uneven surface, reflected sound in transmitted in more than one direction but because of the large surface, most of it gets back to the transducer and the final image is still quite bright. These discussed possibilities are illustrated in figure 2.2.

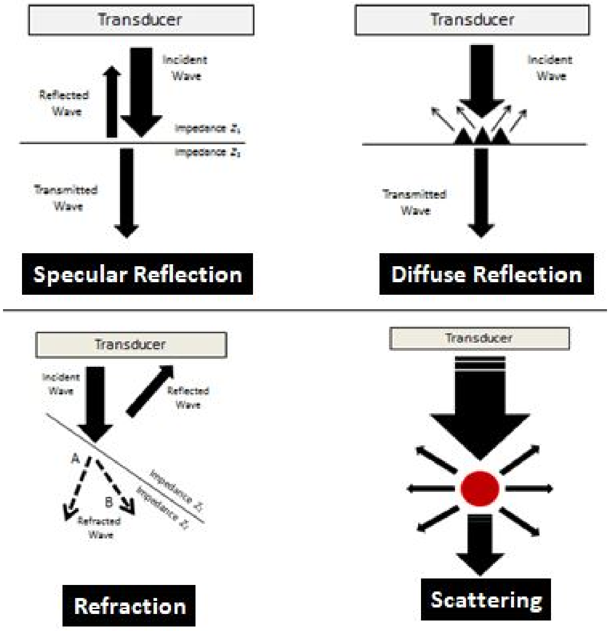


Figure 2.2: Sound waves in an ultrasound [16]

Usually a Doppler ultrasound is also part of an ultrasonography. Doppler effect explains how the frequency of a sound wave is perceived to change relative to movement. This effect is applied in ultrasound to study the blood flow in the subject. Hence, in addition to information such as tissue size, shape and depth even the statistics about the blood flow such as speed, amount per unit time etc can be measured. It is also possible today to convert the sound waves to 3-D images producing 3-D ultrasound.

Ultrasound hence proves to be a real-time, quick and easy method for imaging subjects with zero risks involved. Other imaging techniques such as Magnetics Resonance Imaging (MRI), X-rays and Computed Tomography (CT) have an advantage that they can probe within tissues and produce a clearer result in most cases. However, they require expensive equipment with fixed installation. While X-rays and CT scans use radiation which is pose potential risk of developing cancer [24]. MRI uses very strong magnets and is considered unsafe during pregnancies and in cases where any ferromagnetic objects maybe present; for example a cardiac pacemaker [19].

Therefore, driven by the motivation to reduce the need of such risky imaging, a lot of research has been done in improving ultrasound image quality. Ultrasound elastography measures the mechanical properties of tissues like its stiffness [36]. Contrast Enhanced Ultrasound (CEUS) introduced contrast agents to improve the reliability and quality of ultrasound. The ultrasound data used for analysis in this thesis is from CEUS.

2.2 Contrast Enhanced Ultrasound

The International Contrast Ultrasound Society defines CEUS as an enhanced form of ultrasound imaging that uses biocompatible contrast agents to improve the quality and reliability of ultrasound scans [11].

Conventional ultrasound is used for the examination of the anatomy as a first-line tool for differential diagnosis since it can accurately differentiate cysts or tumors from solid lesions whereas Doppler ultrasound gives information about the blood flow in the subject. However, Doppler ultrasound does not serve the purpose of blood flow inspection in small tissues or microvasculatures. Further, some lesions are inconclusive on conventional gray scale ultrasound, and additional assessment other imaging techniques may be needed for the differential diagnosis. Ultrasound contrast agents overcome these drawbacks of ultrasound.

Contrast agents consist of microbubbles filled with air or gases and this makes them increase the backscatter of ultrasound waves. Currently administered UCAs have low solubility and therefore result in more stable and reliable images even in low acoustic pressure. Since ultrasound is real-time, this property enables dynamic enhancement and promotes effective investigation. Further, the size of the microbubbles is equal to or smaller than the red blood cells. Hence, they are capable of the analysis of both microvascular and microvascular tissues [29].

Based on the type of gas within the microbubbles, contrast agents are classified into two generations. The first-generation ultrasound contrast agent called Levovist was introduced in 1996 and consisted of air within the microbubble shell. The microbubble shell of Levovist allows easy diffusion of air within to the blood pool outside the shell. Furthermore, because of the high solubility of air in blood, the diffused air can easily dissolve into the blood more quickly than preferred [27]. This makes the CEUS unreliable. In order to fix this problem, research was done to make the membrane thicker and to make the air ”dense”. Finally the latter worked out and the second generation of contrast agents were formed. The air inside the microbubbles was replaced by an inert or more slowly diffusing gas such as sulfur hexafluoride or perfluorobutane. SonoVue, Definity , Optison and Sonazoid are some of the second generation contrast agents. Of these agents, SonoVue and Sonazoid (in Japan and Korea) are used for the detection and evaluation of liver lesions [27].

The two generations of contrast agents are well differentialted by the ultrasound metric of Mechanical Index (MI). Mechanical index is a unitless measure of the amplitude of the US wave defined as

$$MI = \frac{P}{\sqrt{f_o}}$$

where P is the peak negative pressure and f_o is the frequency of the wave. MI is used to monitor the possible non-thermal side-effects of contrast agents such as cavitation or tissue degradation. First generation contrast agents have high MI whereas the second generation have a low MI. Low MI enables generation of real-time, stable images and hence the detection and characterization of tumors. Regulations require the MI to be lower than 1.9. First generation agents have an $MI > 0.7$ while the second generation function with an $MI < 0.3$ [27].

Guidelines and good clinical practice recommendations for CEUS [29] define enhancement of a signal as its intensity relative to that of adjacent parenchyma. The signal can be either isoenhancing, hyperenhancing or hypoenhancing when its intensity is equal to, greater than or lower than its neighbor respectively. Most malignant lesions are hypoenhancing while benign lesions are either isoenhancing or hyperenhancing. A lesion can also show sustained or continuous enhancement and on the contrary, a complete absence of enhancement too. ”Wash in” refers to the period before peak enhancement and ”wash out” refers to the period after.

The liver receives 25% - 30% of its blood supply from the hepatic artery and the remaining from the portal vein. This gives rise to three overlapping vascular phases. The arterial phase provides information on based on the blood supply from the hepatic artery. It generally starts 20s after the injection of the contrast agent and lasts for 30-45 seconds. The portal venous phase relates to the blood supply from the porter vein and usually lasts for 2 minutes after the injection. The late phase lasts till the UCA is cleared out of the blood pool and lasts between 4 to 6 minutes. An additional phase called the postvascular or Kpuffer phase , seen with contrast agent Sonazoid, begins 10 minutes after injecting the agent and lasts for an hour or more.

Tang et al. [85] explain the three different ways for administration of CEUS and the following analysis.

◇ Bolus Injections: Administering the contrast agent intravenously ensures that the entire dose enters the general circulation. The agent is distributed throughout the body and finally eliminated by the liver or kidney. When this administration method is used, functional studies are based on the enhancement of ultrasound typically 1-3 minutes after the injection. The analysis can be done both over a large tissue or on a per pixel basis. Enhancements patterns are studied at all the 4 vascular phases. The liver lesions are characterized by comparison of their *time intensity curve* (TIC) with that of normal liver parenchyma during all vascular phases. An example of differentiating TICs is shown in Figure 2.3

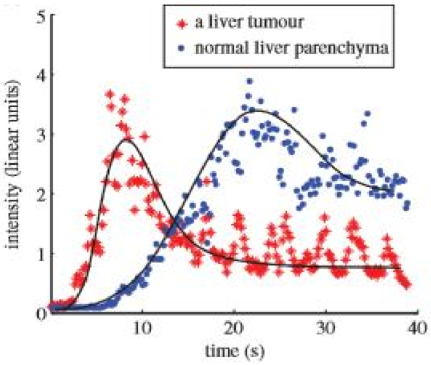


Figure 2.3: Time Intensity curve for a bolus injection in a liver. [85]

Several functional features corresponding to the tissue subject can be determined from the time intensity curve. Some of these are: peak intensity to identify the fractional blood volume and flow, mean transit time and area under the curve. The features used in this thesis are listed in table 4.2 in section 4.2.1.

- ◇ The disruption-replenishment or reperfusion method: This is a two step process where first the contrast agent is bolus injected and once the microbubbles fill the tissue in observation, a series of high intensity pulses is transmitted to the tissue to destroy the bubbles within it. In the second step, the tissue is refilled with microbubbles and the scanner switches to low MI and the refill is monitored with a contrast-specific imaging mode of the ultrasound machine. Unlike the Bolus injection method's TIC, here an exponential curve is seen. The slope corresponds to the velocity of blood flow while the maximum enhancement relates to blood volume. An example of the same is shown in figure 2.4

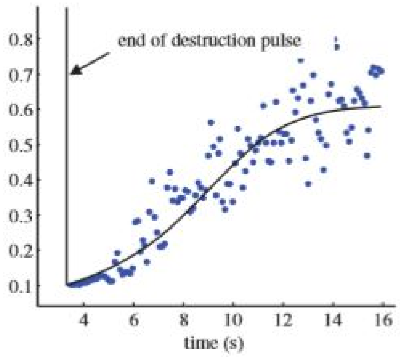


Figure 2.4: Time Intensity curve for a reperfusion. [85]

- ◇ Hepatic vein transit times: Bolus injection of the contrast agent is also used in this method. However instead of TICs, the shortening of the transit time of the contrast agents between the hepatic artery or portal veins and hepatic veins is monitored. The shortening can be a result of a present malignancy. However, this method doesn't identify lesions with high confidence and is rarely used.

CEUS offers notable advantages when compared to Contrast enhanced CT (CECT) or MRI. In addition to being radiation free, affordable and easy, because the arterial phase is short, lasting only for 20-25 seconds, the real-time nature of ultrasound enables capturing it. CT and MRI can not capture real-time information. CEUS has been used to successfully diagnose 98.5% of cases inconclusive on both CECT and conventional grayscale US and led to changes in the treatment plans in 11.6% of these patients [27]. There is no need to predefine scan time points or to perform bolus tracking. The dosage of UCA can be changed based on patient history. Moreover, the excellent tolerance and safety profiles of UCA allow for their repeated administrations in the same session when needed [29].

2.3 Diagnosis of Focal Liver Lesions

Cirrhosis is a condition in which liver cells are irreversibly scarred and is the leading cause of Liver Cancer or Hepatic Cancer. Figure 2.5 shows the liver with cirrhosis and HCC in comparison to a healthy liver.

Focal liver lesions or tumors are of two broad categories.

1. Benign Lesions: These lesions are characterized by a sustained enhancement during the portal venous and the late phases. Further, they show distinct enhancement patterns for

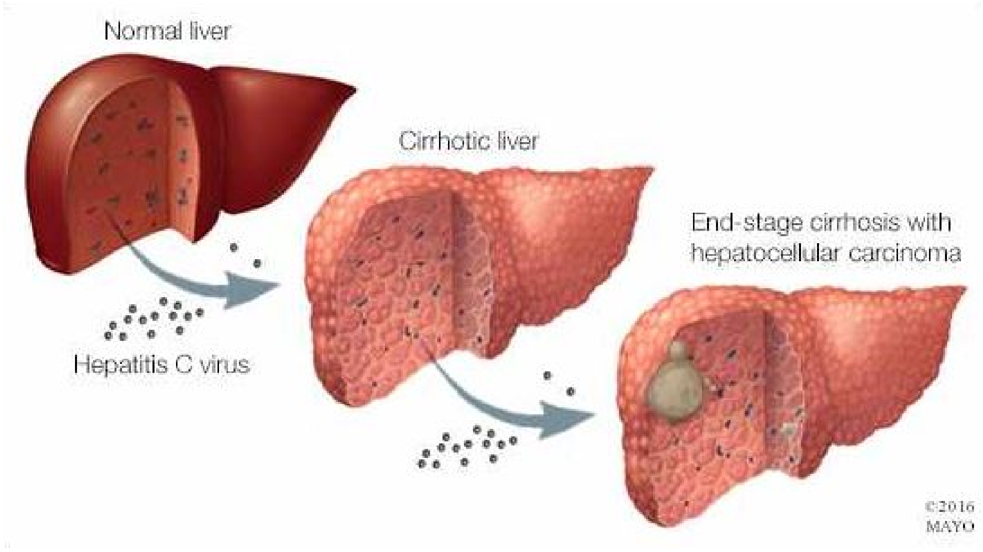


Figure 2.5: Healthy and affected livers [10].

individual kinds of benign lesions in the arterial phase too [21]. They can be present in both cirrhotic and non-cirrhotic livers. These are summarized in Table 2.1

2. Malignant(Cancerous) Lesions: The lesions are characterized by the wash out of microbubbles during the post vascular and late phases. They are hypoenhancing or isoenhancing in these phases. The arterial phase is most important for the detection of malignant lesions : Hepatocellular carcinoma (HCC) and metases as they are hyperenhanced [21]. Although cirrhosis is the leading cause of Malignant Cancers, they can occur even in healthy livers. The enhancement patterns of malignant lesions are shown in Table 2.2 are the standards followed by physicians during diagnosis.

Figure 2.6 A shows the observed arterial phase and B shows the late phases of a lesion in a non-cirrhotic liver. The arterial phase is hypervascular and the enhancement is even throughout the circular lesion. The late phase is isoenhanced - the intensity is equal to the adjacent liver parenchyma. This lesion was diagnosed as focal nodular hyperplasia (FNH) by US-guided biopsy [27]. Notice how spotting the lesion without the annotated arrows is extremely challenging.

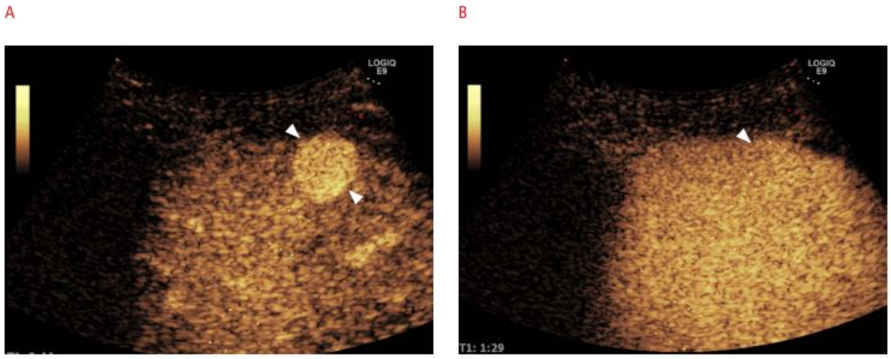


Figure 2.6: The arterial and late phase images of a lesion [27]

Table 2.1: Enhancement pattern of Benign Lesions in cirrhotic and Non-cirrhotic Livers [29].

Lesion	Arterial phase	Portal venous phase	Late phase
A. Noncirrhotic liver			
Hemangioma			
Typical features	Peripheral nodular enhancement	Partial/complete centripetal fill in	Complete enhancement
Additional features	Small lesion: complete, rapid centripetal enhancement		Nonenhancing regions
FNH			
Typical features	Hyperenhancing from the center, complete, early	Hyperenhancing	Iso/hyperenhancing
Additional features	Spoke-wheel arteries Feeding artery	Unenhanced central scar	Unenhanced central scar
Hepatocellular adenoma			
Typical features	Hyperenhancing, complete	Isoenhancing	Isoenhancing
Additional features	Nonenhancing regions	Hyperenhancing Nonenhancing regions	Slightly hypoenhancing Nonenhancing regions
Focal fatty infiltration			
Typical features	Isoenhancing	Isoenhancing	Isoenhancing
Focal fatty sparing			
Typical features	Isoenhancing	Isoenhancing	Isoenhancing
Abscess			
Typical features	Peripheral enhancement, no central enhancement	Hyper-/isoenhancing rim, no central enhancement	Hypoenhancing rim, no central enhancement
Additional features		Hypoenhancing rim	
	Enhanced septa	Enhanced septa	
	Hyperenhanced liver segment	Hyperenhanced liver segment	
Simple cyst			
Typical features	Nonenhancing	Nonenhancing	Nonenhancing
B. Cirrhotic liver			
Regenerative nodule (±dysplastic)			
Typical features (not diagnostic)	Isoenhancing	Isoenhancing	Isoenhancing
Additional features	Hypoenhancing		

Table 2.2: Enhancement pattern of Malignant Lesions in cirrhotic and Non-cirrhotic Livers [29].

Lesion	Arterial phase	Portal venous phase	Late phase
A. Noncirrhotic liver			
Metastasis			
Typical features	Rim-enhancement	Hypoenhancing	Hypo/nonenhancing
Additional features	Complete enhancement	Nonenhancing regions	Nonenhancing regions
	Hyperenhancement		
	Nonenhancing regions		
HCC			
Typical features	Hyperenhancing	Isoenhancing	Hypo/nonenhancing
Additional features	Nonenhancing regions	Nonenhancing regions	Nonenhancing regions
Cholangiocarcinoma			
Typical features	Rim-like hyperenhancement, central hypoenhancement	Hypoenhancing	Nonenhancing
Additional features	Nonenhancing regions	Nonenhancing regions	Nonenhancing regions
	Inhomogeneous		
	Hyperenhancement		
B. Cirrhotic liver			
HCC			
Typical features	Hyperenhancing, complete	Isoenhancing	Hypoenhancing (slightly or moderately)
	Nonenhancing areas (if large)	Nonenhancing regions	
Additional features	Basket pattern, chaotic vessels		Isoenhancing
	Enhancing tumor thrombus		
	Hypo/nonenhancing	Nonenhancing	Nonenhancing

The CEUS images are shown on a dual screen format with low MI B-mode image alongside the contrast-only display. This enables anatomic guidance for small lesions to ensure that the target is kept within the field of view [29]. The data used for this thesis is collected from a machine with such imaging mode. A frame from a CEUS video of a patient is shown in figure 2.7.

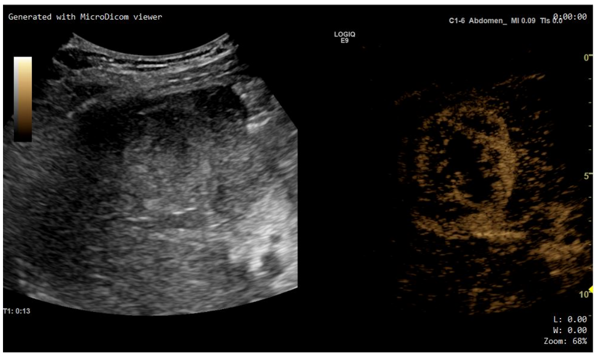

Generated with MicroDicom viewer
C1-6 Abdomen_ MI 0.09 TIs 0.0
0:00:00
LOGIQ
E9
0
5
10
T1: 0:13
L: 0.00
W: 0.00
Zoom: 68%

Figure 2.7: A dual display of B-mode US and CEUS

However, a difficulty with the split screen method is that the low MI is used for both B-mode and CEUS panels. This means that the gray scale display is noisy. So, the smaller and low contrast lesions may be difficult to image. On some scanners, conventional and CEUS images are not split onto two different screens but overlaid with different color scales [29]. Section 3.3 describes the noise handling used in this study.

Two major hurdles in detecting HCC particularly are that:

◇ HCC shows hypervascularity in the short arterial phase and hypovascularity starting from wash out into the post vascular and late phases.

◇ The differentiability of the HCC with the adjacent liver parenchyma determine detection accuracy. Moderately differentiated HCC generally show classic enhancement features, while well and poorly differentiated tumors account for most atypical variations [43].

78% of well differentiated HCC showed atypical behavior [27]. Hence, identifying a HCC lesion is technically challenging.

2.4 Digital Imaging and Communications in Medicine (DICOM)

The American College of Radiology (ACR) and the National Electrical Manufacturers Association (NEMA) together came up with the a standard for medical data communication in 1983. This was needed to enable medical imaging centers to efficiently share data among them. The third version of this standard, in 1993 was named ”DICOM”. Today, DICOM is used extensively in hospitals and ensures the interoperability of producing, processing and storing patient data.

Information Object Definition (IOD) is an object-oriented representation of the real world in DICOM. A normalized IOD represents a single real-world entity like patient personal details.

The Composite IOD comprises of multiple entities together. Such as ultrasound scans of the patient along with personal data. The relevant IODs extracted from the DICOM data used for this study are:

1. Rows: Defines the rows of the frame read.

2. Columns: Defines the columns of the frame read.

3. Pixel Data: A DICOM data set (Figure 2.8) consists one special element containing the image pixel data called ”Pixel Data”. A single DICOM dataset can have only one element containing pixel data. However, the element may contain multiple “frames”, allowing storage of multi-frame data like videos [6].

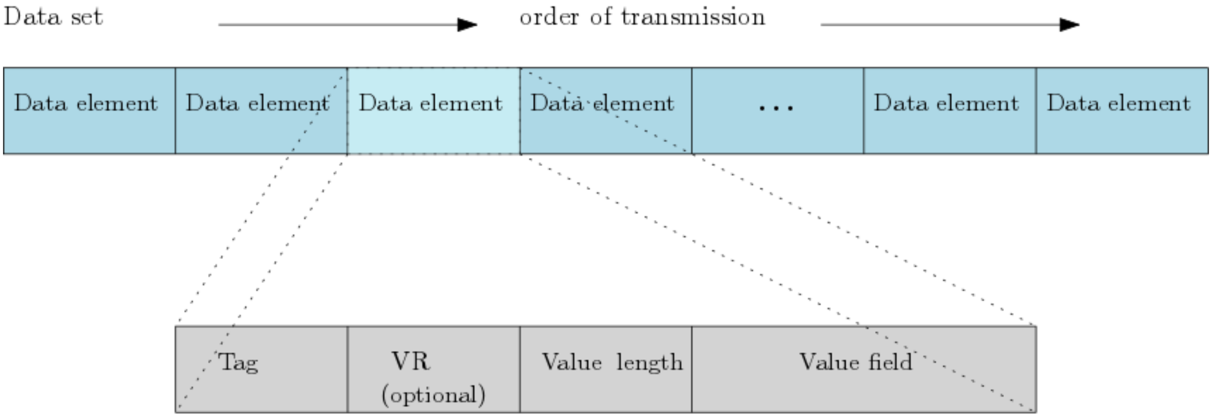


Figure 2.8: A depiction of DICOM dataset [5]

The DICOM data format groups information into data sets. That means that a file of a ultrasound scan image, for instance, actually contains the patient ID within the file so that the image can never be separated from this information by mistake. Figure 2.8 shows the structure of data elements that make up the dataset. The data elements in DICOM’s latest revision are ordered according to the Little Endian byte ordering. Hence, in a characters are encoded in the order of occurrence i.e. from left to right. There are various different formats for the data element prescribed by the standard. However, each have three common fields:

- ◇ Tag: A 16-bit unique identifier representing the group number, indicating the function of the IOD and the element number.

- ◇ Value Length: The length of the value of the element.

- ◇ Value: The value of the element

The pixel data conveyed in the Pixel Data IOD can be either in the Native or uncompressed format, or in a compressed format (not supported by the DICOM standard). In the native format, Pixel Cells are encoded as the direct concatenation of the bits of each Pixel Cell in little endian format. That is, it is ordered from the least significant to the most significant bit. Pixel Cell is the container for a single Pixel Sample Value-a value associated with an individual pixel. The number of bits that make up each Pixel Cell is defined by the ”Bits Allocated” data element Value [7]. In the data used for this study, the Bits Allocated is 8. Pixel Sample Value is associated with an individual pixel. However, a pixel can have multiple Pixel Sample Values. In the data used for this study, every pixel has 3 sample values associated with each. The values correspond to the red, green and blue light components of the pixel respectively. The red, green and blue components are added together to reproduce a color seen in the final image.

The RGB triplet (r,g,b) per pixel can be mapped to a three dimensional coordinate system. This approach is advantageous when the colors of two pixels have to be compared. Their smaller their Eucledian distance, the more similar they are.

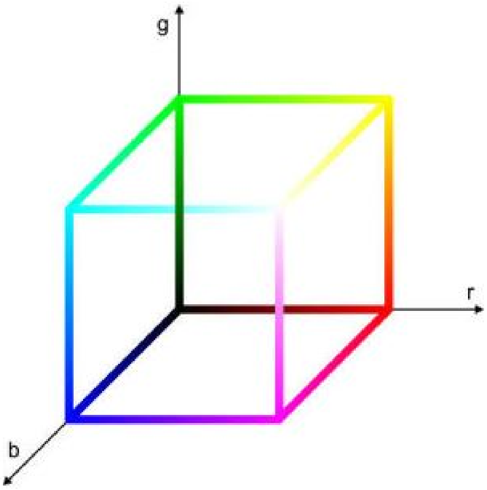


Figure 2.9: The RGB color space [14]

2.5 JPEG

The Joint Photographic Experts Group is the joint committee formed by International Standardization Organization (ISO) and the International Electrotechnical Commission (IEC) in 1992. The committee created the image compression standard - JPEG, in 1992. In this thesis, the output of the CAD system, that is, the resulting images marked with identified tissues are compressed using jpeg.

A survey conducted to compare various lossless compressing techniques for medical images [18], found JPEG-LS to rank highest based on compression speed and compression ratio (Figure 2.10). In this study, the Lossless JPEG compression technique with maximum compression quality and minimum loss is chosen.

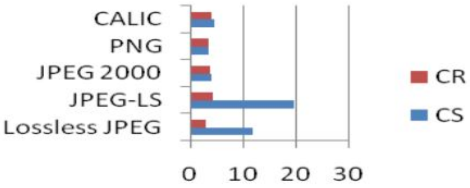


Figure 2.10: Comparison of Compression Rate(CR) and Compression Speed(CS) of various compression algorithms [18]

The measure comparison ratio (CR) , defined as

$$CR = \frac{Original\ Image\ Size}{Comperessed\ Image\ Size}$$

can be set for the compression. When converting the DICOM information to JPEG, we perform the compression at a CR of 1. Hence, no image data is lost during conversion.

2.6 Artificial Intelligence

Artificial Intelligence(AI) or Machine Intelligence is defined as the study of ”*intelligent agents*”- any device, machine or program that perceives its environment and takes actions that maximize its probability of successfully achieving its goals. AI often revolves around the use of algorithms or sets of unambiguous instructions that a mechanical computer can execute. The goal of *learning* from its environment can be accomplished by such algorithms using either repetitive trial and error or a cycle of trial-error-feedback called reinforcement learning. The algorithms can be of 3 broad types based on their design.

1. Symbolic: These task-specific algorithms, widely used from mid-1950s until the late 1980s were made up of production rules consisting of extensive if-then logic.

2. Machine Learning: These algorithms are generally of statistical nature and concentrate heavily on learning data representations as opposed to task-specific algorithms.

3. Cognitive: These algorithms are an integration of symbolic and machine learning based algorithms aimed to build intelligent agents with human like capabilities of performing a variety of tasks, like decision making, problem solving, planning, and natural language understanding, by encoding, using and learning all types of knowledge.

2.6.1 Machine Learning

Tom M. Mitchell [58] defines learning in machines as follows- ”A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E.” A basic machine learning system takes inputs and identifies feature vectors that describe the input. It ”learns” based on these features and creates a model. This model can now be used on a previously unseen input to produce a required output.

Machine learning tasks can be broadly of two categories: *classification* and *regression*. Classification divides input data into two or more discrete categories or classes. Regression is related to forecasting or predicting a continuous output.

Additionally, based on the whether the learning system leverages available user feedback, machine learning systems are of two broad types:

1. Supervised Learning: These algorithms map inputs to corresponding outputs based on given input-output mappings. Predictive functions are learnt from labelled training examples. Both classification and regression tasks are typically tackled as supervised learning problems.

2. Unsupervised Learning: The algorithms are used when the data available is unlabelled. Functions have to *infer* the relationships in the data. In the context of unsupervised learning, grouping of similar data together is termed as *clustering*.

Feature generation and selection is an important if not the most important part of machine learning. The output is heavily dependent on the quality and quantity of features used. Models trained with a large number of features may more often than not suffer from the curse of dimensionality. On the other hand, those trained with too few features might not be able to differentiate all inputs seen in a good enough way. A subset of machine learning algorithms that automate the process of feature generation and selection is *Deep Learning*. They use layers of neural networks to learn data relationships and also features from them.

3

Data Pre-Processing and Exploration

This chapter discusses the properties of the collected data source. Details of the size and attributes of the data are described in sections 3.1. The setup of the data for the experiments carried out is explained in section 3.2. The steps taken to clean the data are enumerated in section 3.3. In addition to the preprocessing details, in section 3.4, we compare the brightness and intensities of the CEUS pixels of various tissues statistically to test their similarity. Also, the CEUS brightness and intensity of a single tissue are compared.

3.1 Data Exploration

Dual mode contrast enhanced DICOM videos of the liver of 11 patients were collected from University Hospital, Munster from the ultrasound machine - GE Healthcare Logiq E9 and using SonoVue as contrast agent. These accounted for 3.5 GB of video data and translated to 5475 frames in total.

The DICOM video was cut into frames using the opensource DICOM viewer MicroDicom [13]. It was observed that additional unusable information was present around the US and CEUS images in the frames extracted. These would ideally represent patient details but they have been removed by the hospital so as to not disclose personal information. They were trimmed out without tampering the content of interest. Further, it was observed that most interesting information was present in the vertical center of the frames for the majority of the samples. The CEUS side of the frames cover the same tissues as their B-mode US counterparts but are translated right horizontally. For every pixel, the DICOM format carries its RGB values.

3.2 Data Setup

The experiments conducted in this thesis can be categorized into three based on the methods used.

1. Time-series based: Out of the total 11 videos, 2 videos were found to have very well defined contrast enhanced portion. These videos, amounting to a total of 623 frames were the focus for this part of the study.

2. Deep Learning based: Videos of 4 patients showing different levels of CEUS side images were chosen out of the total 11 videos for this part of the study. The four videos amounted to 1030 frames.

3. Preliminary US brightness based: For the videos with negligent CEUS components, a separate pipeline was built using only the US data. Since this is not the main focus of the study, we worked with only single patient data to verify this approach.

3.3 Data Pre-processing

Two pre-processing steps were applied to the frames in this study.

- ◇ **Smoothening** : During the scan the physicians move around the transducer quite a bit and exert varied amount of pressure on it while doing so. This is done to capture the image of the target organ and its surroundings from various positions. However, it induces noise in the video due to shaking. This was observed in all the videos and hence a mechanism to minimize the shaking was applied globally. *Smoothening* was used to counter the disturbances caused by shaking. Figure 3.1 illustrates this on a single frame. It is important to highlight here that the smoothening was done over sets of frames belonging to a single video while preserving their order. The pixel values were averaged over a selected selected number of frames or "window size". We conducted our experiments with a window sizes of 5 in this study.

- ◇ **Median Filtering** : Due to random fluctuations of the sound signals, ultrasound images generally suffer from additive noises like gaussian noise, speckle noise and salt and pepper noise. Median filtering has proven to be a good noise reducing technique that also preserves the underlying image features like edges for instance [66]. The RGB vales of a pixel are the median of the RGB values of its *"neighbours'*. In this study we perform a $7x7$ window filtering. That is, a $7x7$ rectangle with current pixel at its center is considered the neighbourhood of the pixel. An example of the resulting transformation is shown in figure 3.2. However, since we expect the deep learning algorithms to generate features based on neighbouring information among others features, we decided to not apply median filtering for frames in the deep learning based experiment.

3.4 Time-Intensity and Time-Brightness Curves

It is widely accepted that liver lesions are characterized by comparison of their *time intensity curve* (TIC) [85]. In this thesis we statistically examine the above. The analysis is conducted in two steps. First, we compare the CEUS brightness and intensity of two tissues in a single video. In the second stage of the analysis, we compare the 2 tissues within CEUS videos of 2 patients.

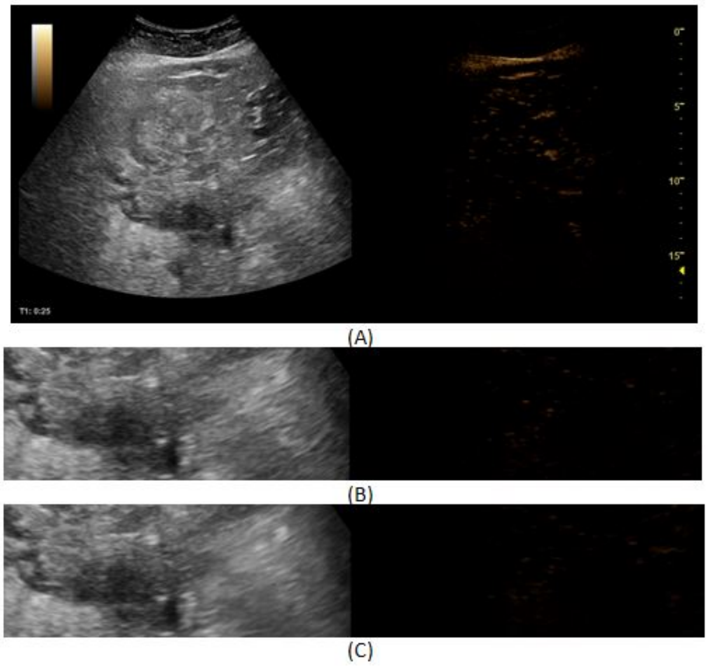


Figure 3.1: (A) Original frame as viewed in MicroDicom viewer. (B) Selected center portion of the frame. (C) Smoothened selection with a window of size 5.

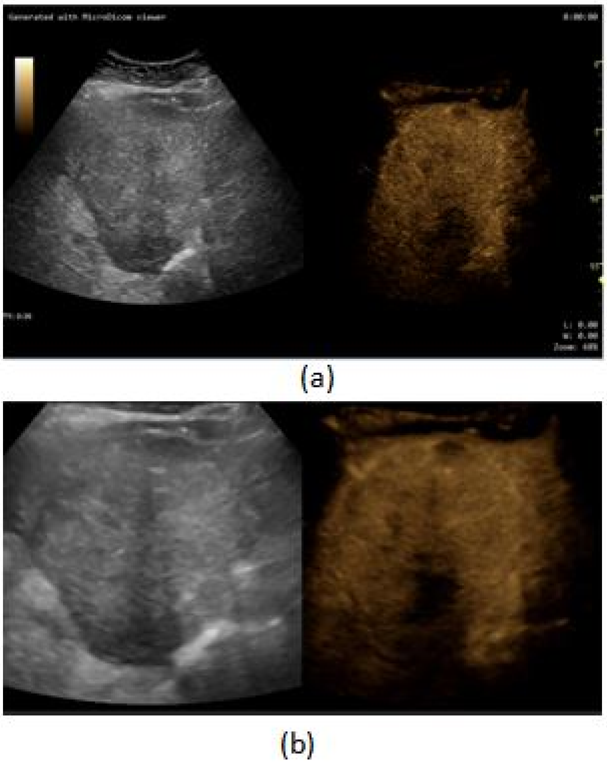


Figure 3.2: (a) Original frame as viewed in MicroDicom viewer. (b) Median filtered (window size=7) and cropped frame.

3.4.1 CEUS Intensity vs Brightness

Studies concentrate on the intensity of CEUS to study the TIC. We compare the shapes of the CEUS brightness and intensity curves in figures 3.3 and 3.4 and find them to be similar. Section 3.4.3, statistically validates this assumption made. However, in the case of tumors, the brightness is more pronounced as seen in figure 3.4. The curve is not smooth due to the shaking in the Sonography process. Smoothening and median filtering reduce noise but do not eradicate it. Also, we observe that the sonography video available does not have complete data until the contrast agent is washed out of the blood stream.

We will use the CEUS brightness henceforth. This approach is chosen because the brightness calculated as

$$Brightness = 0.299 * R + 0.587 * G + 0.114 * B$$

weighs the red(R) and green(G) components of the pixels and they are the ones absorbed most by the human eye. Whereas the intensity is the average of the three color components.

$$Intensity = \frac{R+G+B}{3}$$

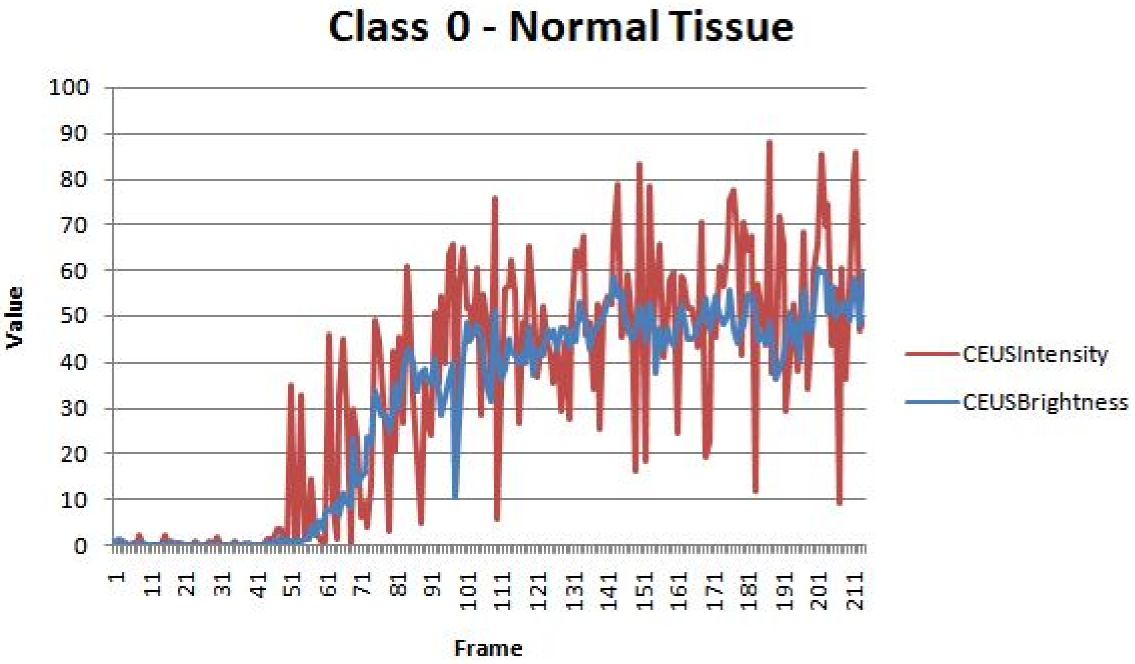


Figure 3.3: Brightness and Intensity curve of CEUS pixels in a normal tissue. The intensity is more pronounced.

3.4.2 Two sample tests

Several statistical tests are available to compare the similarity between two samples.

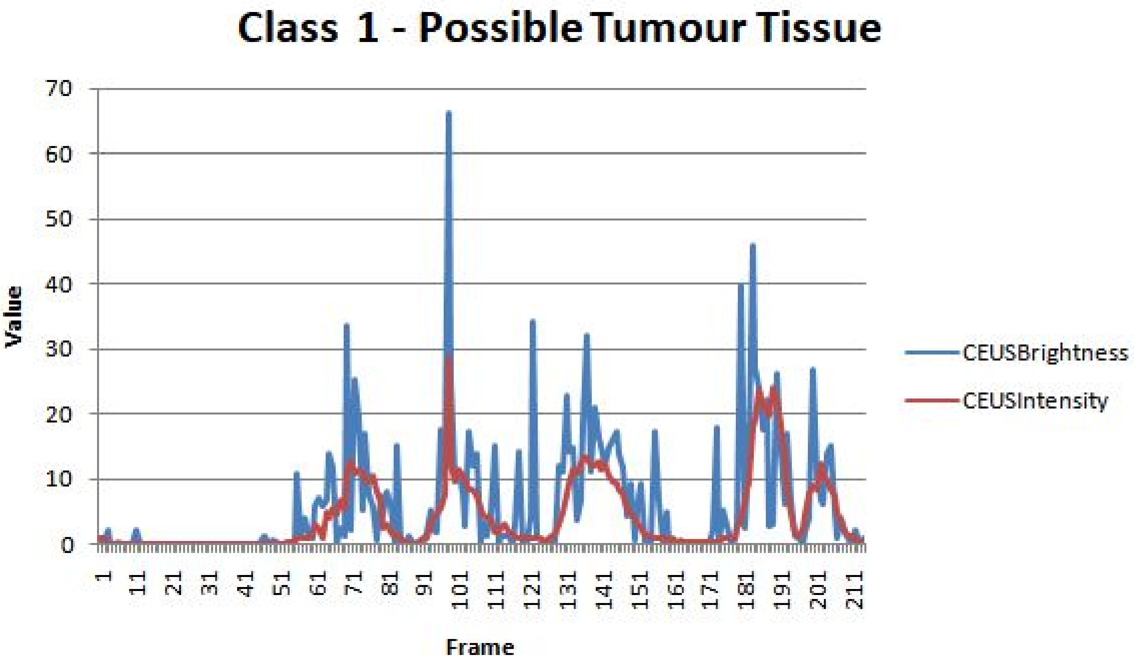


Figure 3.4: Brightness and Intensity curve of CEUS pixels in a tumour tissue. The brightness is more pronounced.

Two sample T-test

The two sample T-test checks if the means of two independent samples are equal. The null hypothesis of the test is

$$H_0 : \mu_1 = \mu_2$$

The test statistic T is given by

$$T = \frac{\bar{y_1} - \bar{y_2}}{s_p(\frac{1}{n_1} + \frac{1}{n_2})}$$

where y_i is the average of group i , n_i is the number of items in group i and under the assumption of equal variances, s_p is the pooled variance given by

$$s_p = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

where s_i is the standard deviation of group i. Whereas, under the assumption of unequal variances, the test statistic is given by

$$T = \frac{\bar{y_1} - \bar{y_2}}{(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2})}$$

The null hypothesis is rejected if

$$|T| > t_{(1-\alpha/2),df}$$

where $t_{(1-\alpha/2),df}$ is the critical value of the t-distribution with degree of freedom df and confidence $1 - \alpha/2$. The degrees of freedom df is $n_1 + n_2 - 2$ if the two groups are homoscedastic. If not,

$$df = \frac{(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2})^2}{\frac{s_1^4}{(n_1-1)n_1^2} + \frac{s_2^4}{(n_2-1)n_2^2}}$$

However, the two sample t-test assumes that the underlying groups are normal. Hence, we first test the normality of our data.

Test for Normality

Shapiro Wilk is a powerful test for normality but it is sensitive to ties in data. And, in our case, there are several times pixels might have the same intensity or brightness. Hence, we chose the Anderson Darling test which works well for data with ties and also the QQ plot to visualize the sample quantiles against the theoretical quantiles from a standard normal distribution to check for normality.

The QQ plots for the intensity and brightness of CEUS pixels of a class from a video with 219 frames are shown in figures 3.5 and 3.6. They seem to deviate from normality. The Anderson Darling test produces a p value of $< 2.2e-16$ for both CEUS intensity and brightness of labelled class 0. Hence, it can be concluded that the CEUS intensity and brightness of labelled class 0 are both not normal.

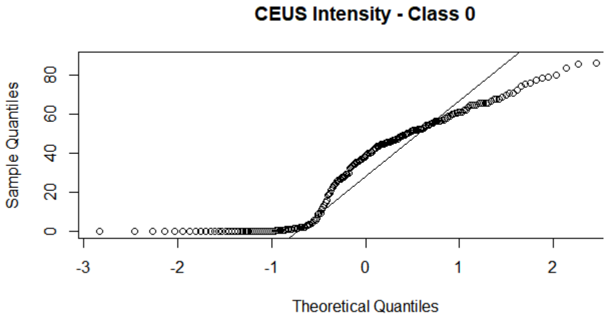


Figure 3.5: QQ plot for the CEUS pixel intensity.

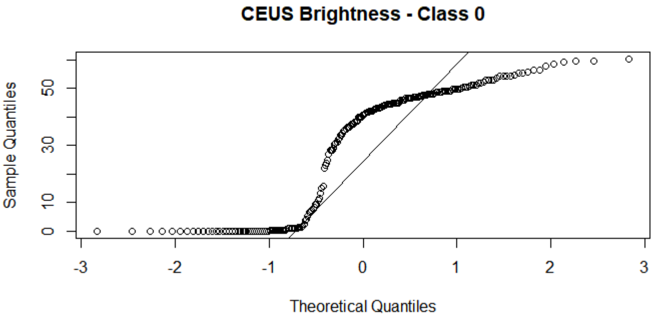


Figure 3.6: QQ plot for the CEUS pixel brightness.

Whereas, the QQ plot of the brightness of corresponding US pixels of the same class shown in figure 3.7 suggests normality. This is also supported by the Anderson Darling test($\alpha = 0.05$) with a p value of 0.159.

For pixels labelled class 1, denoting tumour tissues, a similar observation was made.The CEUS brightness and intensities did not show normality while the US brightness did not reject normality.

Hence, we used the two-sample t-test to compare the equality of means for only the US brightness of the pixels of two classes - class 0 and class 1 in a video with 219 frames. We used R to run the statistical test and the null hypothesis was rejected with a p value of $< 2.2e-16$. That is there is statistical significance that the US brightness of the two classes of tissues have different means.

Since the CEUS intensities and brightnesses of the two class do not show normality, we use a non-parametric test - Wilcoxon rank sum test, to compare them.

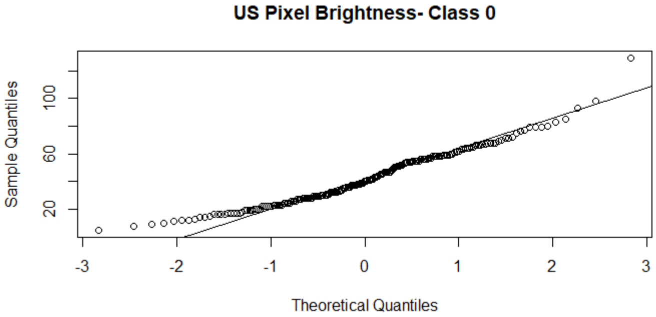


Figure 3.7: QQ plot for the US pixel brightness of labelled tissue class 0.

3.4.3 Wilcoxon Rank Sum Test

This test is often described as the test for equal medians. The null hypothesis of the Wilcoxon rank sum test is $H_0 : m_1 = m_2$ where m_i is the median of sample i. Wilcoxon rank sum test is non-parametric and hence, does not require the groups being compared to show normality. Let $y_{i1}, y_{i2}, y_{i3}, ..., y_{in_i}$ be the data from group i. Rank $r(y_{ij})$ is defined as the rank of y_{ij} in the complete data set $y_{11}, y_{12}, y_{13}, ..., y_{1n_1}, y_{21}, y_{22}, y_{23}, ..., y_{2n_2}$. The sum of ranks for the two groups is calculated as

$$S_i = \sum_{j=1}^{n_i} r(y_{ij}) \quad \forall i = 1,2$$

The test static of the Wilcoxon rank sum test W under large sample approximation is

$$W = \frac{s - \mu(S)}{\sigma(S)} \qquad ,S = S_1 \quad or \quad S = S_2$$

where,

$$\mu(S_i) = \frac{n_i(n_1 + n_2 + 1)}{2}$$

$$\sigma^2(S_i) = \frac{n_1 n_2(n_1 + n_2 + 1}{12}$$

The test static is compared to the quantiles of the standard normal distribution. As shown in the previous section, the CEUS intensity and brightness of tissues reject normality. Hence, we use the Wilcoxon rank sum test to compare them. In both cases, the test results in a p value $< 2.2e - 16$. This implies the null hypothesis can not be accepted. That is, there is statistically significant evidence that CEUS intensities and brightnesses of the pixels of the two classes are different.

Comparison of CEUS Brightness and Intensity

We also use the Wilcoxon rank sum test to verify the assumption made in section 3.4.1. We test for the equality of the medians of the CEUS Brightness and Intensity of tissues labelled class 0 (data points plotted in figures 3.3 and 3.4). The equality of the medians can not be rejected with a p value of 0.09. Hence, the CEUS brightness and intensity curves are statistically similar.

Therefore, it is verified that the two classes of tissues are statistically different both in terms of their US and CEUS measurements. Further, the CEUS brightness is more pronounced in magnitude than the intensity but both show the same statistical behavior when compared between classes. We choose to use the CEUS brightness for the classification.

3.4.4 Multiple Sample Homogeneity Test

In the previous sections, we compared classes of tissues within a single video and found that there are statistical differences between them. In this section, we test if various classes in multiple patient videos are distinguishable. Two sample tests should not be repeatedly applied in combinations to compare multiple samples. To compare the means of multiple samples, the analysis of variance (ANOVA) test is adopted. ANOVA compares the variation between groups to the variation within groups. Figure 3.8 shows the box-plots of the CEUS brightness of 4 classes of tissues from two patient videos. The videos chosen had 414 and 219 frames each. To ensure our samples are unbiased, we adopt truncation and use only the first 219 frames of the longer video. Visually, the classes are distinguishable. The results of ANOVA test echo this observation. With a p value of $< 2e-16$, the null hypothesis that the means of the samples are equal, can be rejected.

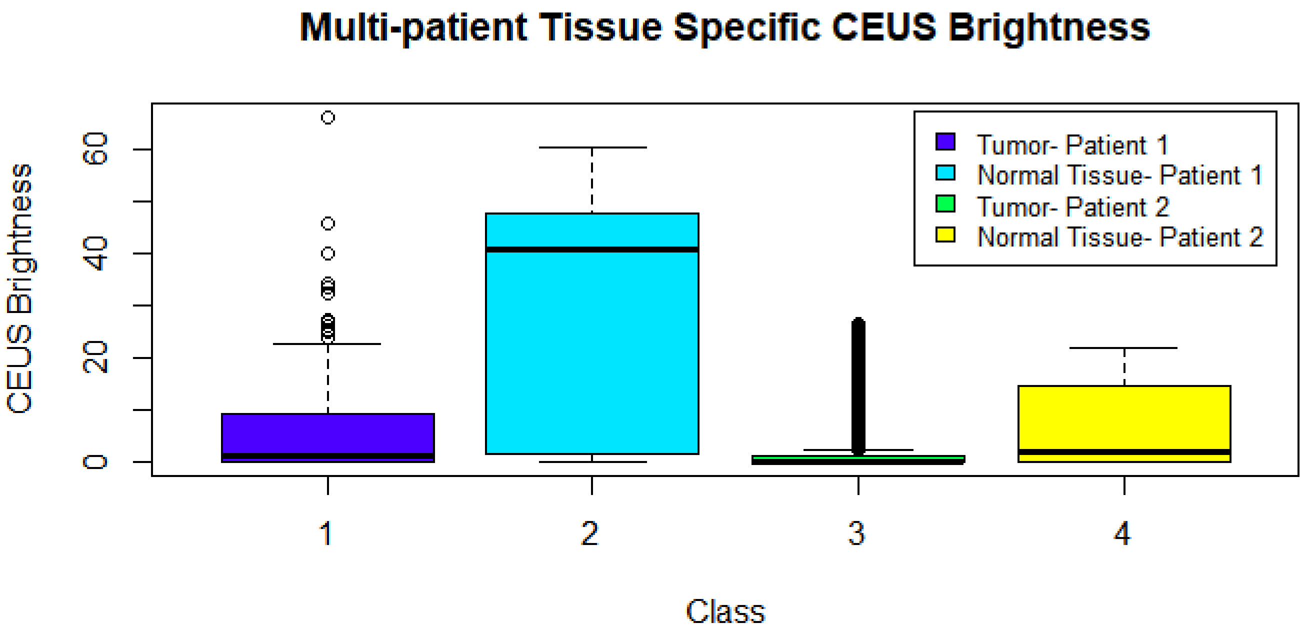


Figure 3.8: Boxplot of the CEUS brightness observed in two classes in two different patients.

Taking a step further, we also test if the same tissue class is identical between patients. We compare the normal tissue (labelled class 1) classes and tumor tissue (labelled class 0) classes, each, between the two patients. The equality of means of the CEUS brightness of the tumor classes between the two videos is rejected with a p value of $2.27e-05$. The mean CEUS brightness of the normal tissues are also not the same between the two patients ($p < 2e-16$). Although, these results obtained by comparing merely two patients are not sufficient to generalize but they echo the patient based factors explained by Dietrich et al. in [32]. To validate the effects of this, in our experiments, we train and compare models with both single patient data and combination of multiple patient data.

4
Tissue Classification

This chapter elaborates the main contributions of the thesis. It explains the algorithms applied on the dataset after the preprocessing described in the previous chapter. Section 4.1 presents an approach applied to videos with weak CEUS components. Following sections 4.2 and 4.3 provide in depth details on the artificial intelligence approaches adopted for CEUS videos in this study.

4.1 Sonographs with weak CEUS components

It was seen that 8 out the total 11 videos did not show a prominent CEUS side image. This maybe due to presence of minuscule benign tumors. Hence, we used only the US side of the frames to identify tissues in such videos. This is not the most important contribution of this thesis but it allows tissue recognition in the absence of good quality CEUS and hence is important nevertheless.

At this point of the thesis, labelled data was not made available by the hospital. Hence, two approaches were tried to label the data. First, the US brightness values of pixels were split into three equal sized bins and the pixels were put in on of the bins based on their US brightness values. A pixel's bin number was the class assigned to the pixel. One tissue of interest can be clearly seen with the lowest US brightness (the darkest region) in the 'Unlabelled smoothened original' image in figure 4.1. As a second approach, we marked rectangular region of interest per frame of the video which is also shown in figure 4.1. It shows two classes marked by green and red.

We used a PML classification pipeline for the classification. As explained in section 6.2.1, pixel-based machine learning is widely being used in the medical domain. For every frame f of a video consisting on n pixels, every pixel $p_i \quad \forall \quad i \in [1,n]$ is labelled. The algorithms used in this thesis are from the WEKA toolkit integrated with Java. Features to identify individual pixels are determined and instances are built for every pixel with these features. Features are chosen based on the data labelling method, the pixels position, neighborhood and its color channels. Using Canny's edge detection, edges are found within the frame. These are important as they mark tissue boundaries. A binary feature to determine a pixel's position on an edge is created. The median of these features in the 3x3 neighborhood of the current pixel is also

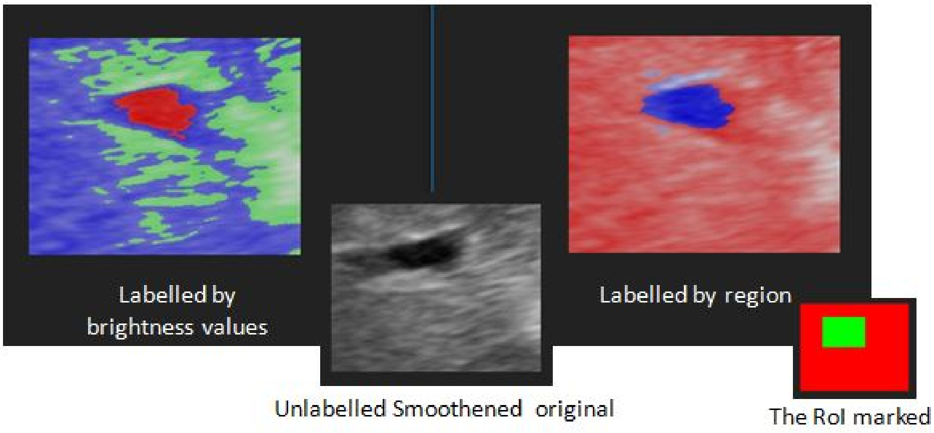


Figure 4.1: The result of classification using Hoeffding Tree classifier with different approaches to labelling.

considered. Moreover, the features chosen are dependent on the method of labelling. For labelling based on the pixel brightness values, the 8 features are described in table 4.1.

Feature description	Number of features	Type
Pixel color components: red, green, blue.	3	Numerical
Pixel's position on an edge	1	Binary
The median color components(red, green, blue) of its 9 neighbors.	3	Numerical
Sum of neighbors on edge.	1	Numerical

Table 4.1: Pixel features when data labelled with US brightness values.

On the other hand, when the labelling was done solely based on the pixel's position by drawing rectangular RoIs, brightness was used as a measure instead of the individual color components. This reduced the number of features to 4: pixel US brightness(numerical), pixel on edge (binary), median pixel neighborhood brightness (numerical) and sum of pixel neighbors on edge (numerical).

Apart from the above features, a nominal feature *'class'* denotes the class in which the pixel belongs. These feature are used as attributes to form instances required by WEKA classifiers.

The frames are split into 70-30 training and testing sets. We test the performance of classifiers: Naive Bayes (refer section 4.2.1.1) and Hoeffding Tree (refer section 4.2.1.3). The outcome of the Hoeffding Tree classifier using both labelling methods is shown is figure 4.1. In both cases the region of interest was identified. In the first case, with 3 classes, additionally probable fat tissues are also identified as a tissue marked by green on the left of figure 4.1 .

4.2 A Time-series Approach

The main idea behind the use of contrast agents for ultrasound is that they result in distinguishable time-intensity curves for different tissues [85]. This has been statistically verified in section 3.4. We leverage the differences in the time-brightness curves of tissues and approach the tissue classification as a time-series classification problem.

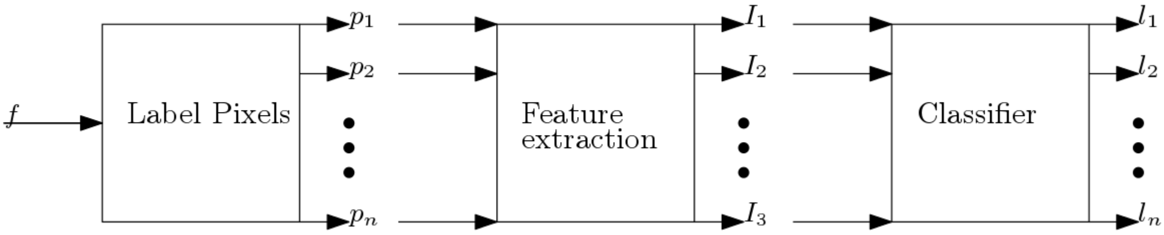


Figure 4.2: Per pixel classification pipeline for US.

Regions of interest are identified for every frame of a video. This information has been collected from the physicians. A single RoI can not be used for all frames of a video because its position might change throughout the video as the radiologist moves the transducer. An open source image annotation tool [4], was used to annotate the frames of the video. Rectangular boxes are used to mark the RoIs. An example is shown in figure 4.3.

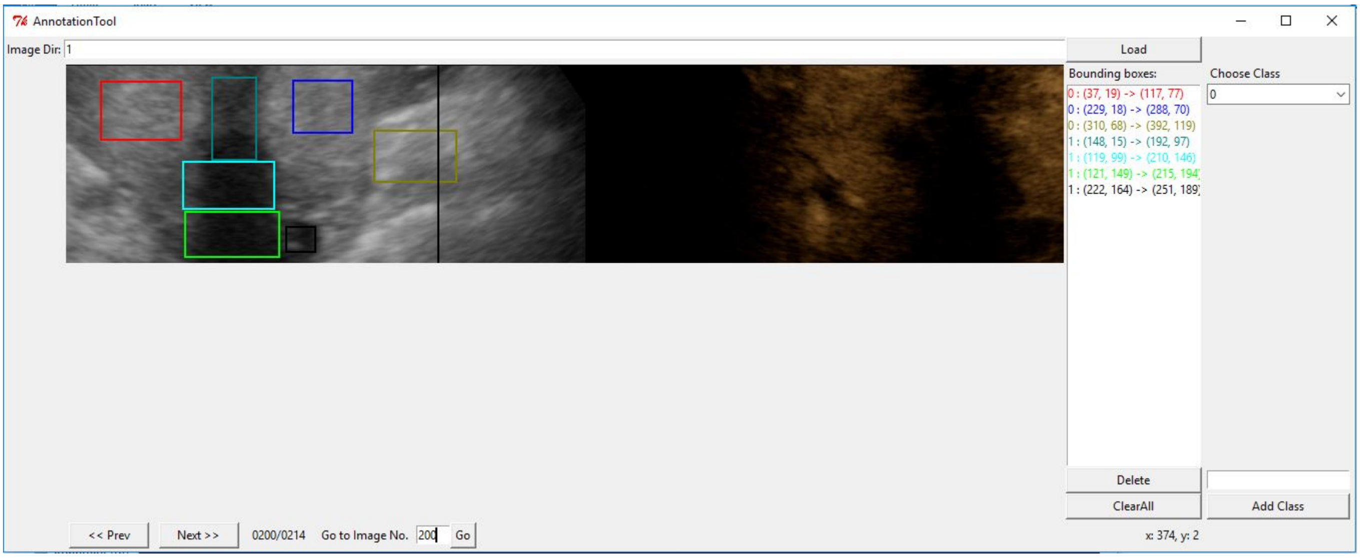


Image Dir: 1

Bounding boxes:

0 : (37, 19) -> (117, 77)
0 : (229, 18) -> (288, 70)
0 : (310, 68) -> (392, 119)
1 : (148, 15) -> (192, 97)
1 : (119, 99) -> (210, 146)
1 : (121, 149) -> (215, 194)
1 : (222, 164) -> (251, 189)

Choose Class: 0

0200/0214 Go to Image No. 200

x: 374, y: 2

Figure 4.3: Example of a frame annotated with 7 rectangles for two classes - 0,1

A supervised learning approach is adopted and the data is first labelled. Every frame in a video is labelled with classes corresponding to RoIs present. For every class in a frame, the pixel with the median CEUS brightness is chosen as the representative for that class in that frame. Hence, irrespective of the size of the RoI in a frame, every class is equally represented in a frame. In this manner, at the end of a video, a frame-brightness curve is created. The frame corresponds to time.

While training, only representative pixels of each RoI marked are looked at. Once a classifier is trained, it is used to classify each pixel in every frame of the video. We first train a model per video for two patients. Next, we combine the data from the two patients and train the models on the combined data. We want to point out that the classifier trained only on one video is not limited to 'memorization'. The learning is generalized over pixels. In order to test the models, videos are converted into frames and passed as input to the model which classifies each pixel of every frame resulting in a classified output image. Finally, the output frames are stitched together to result in a video with tissues identified and color annotated. The system design for a video with two tissues identified using structural feature extraction is shown in figure 4.4.

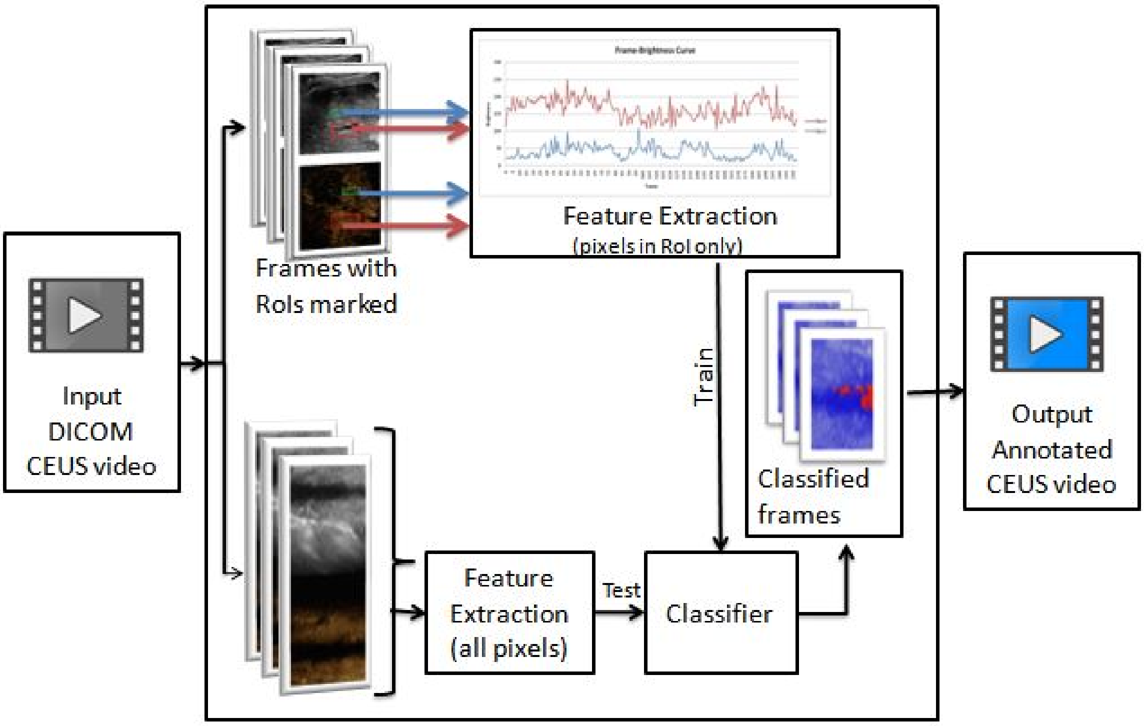


Figure 4.4: System design for a video with two tissues of interest using structural feature extraction.

The system built has a learning phase and a prediction phase. In the former, the time-series for the CEUS brightness of labelled tissues are fed to the model for learning. If there are *n* labelled tissues, a classifier with *n* classes is the result of this phase. In the latter phase, as the video proceeds, a time-series for the CEUS brightness for *all pixels* is formed. These are then classified by the trained classifier into one of the *n* classes.

4.2.1 Structural Feature Extraction

Although Fourier and wavelet transformation are heavily used for feature extraction, owing to the relative small size of data, we chose the structural feature extraction method like Wiens et al. [90]. A bag of features that identify the current state, global state and fluctuation in the time series is prepared. The features used are listed in table 4.2. Feature 1 identifies time or here, the number of frames. Features 2-4 are averages that characterize the series. Linear and Quadratic weighted averages weigh more importance to the most recent observations. Features 6-8 and 19 are indicators for the amount of fluctuation in the observations. Features 5 and 9 gather information about the most recent states the pixel's CEUS brightness. Features 10-13 summarize information regarding global maxima and minima in the series. Key differentiating metrics like the time to peak and time to decay are captured here. Features 14-16 describe the distribution of the brightnesses near the mean. Features 17 and 18 identify micro-clusters of similar pixels above and below the mean.

Feature	Description	Calculation
1	Number of frame/ Length of time-series	f
2	Average brightness	$\mu = \frac{1}{f}\sum_{i=1}^{f} O_i$
3	Linear weighted average brightness	$\frac{2}{f(f+1)}\sum_{i=1}^{f} iO_i$
4	Quadratic weighted average brightness	$\frac{6}{f(f+1)(2f+1)}\sum_{i=1}^{f} i^2 O_i$
5	Current CEUS Brightness	O_f
6	Standard deviation of brightness	σ
7	Average absolute change in brightness between frames	$\frac{1}{f}\sum_{i=1}^{f-1} \lvert O_i - O_{i+1} \rvert$
8	Average absolute change in derivatives of brightness between frames	$\frac{1}{f}\sum_{i=1}^{f-2} \lvert O'_i - O'_{i+1} \rvert$
9	Sum of the brightness seen in the last 3 frames	$\sum_{i=f-3}^{f} O_i$
10	Time to peak/ Frame of maximum brightness	$\frac{1}{f} argmaxO_i$
11	Peak/ maximum brightness	$\max_i O_i$
12	Time to decay/ Frame of minimum brightness	$\frac{1}{f} argminO_i$
13	Minimum brightness	$\min_i O_i$
14	Number of pixels with brightness above mean	$\sum_{O_i>\mu} 1$
15	Number of pixels with brightness below mean	$\sum_{O_i<\mu} 1$
16	Ratio of below mean to above mean	$(\sum_{O_i<\mu} 1)/(\sum_{O_i>\mu} 1)$
17	Longest streak of brightness above mean	
18	Longest streak of brightness below mean	
19	Variance	σ^2

Table 4.2: Features of the time-series: CEUS brightness-frame curve with observation vector $O = [O_1, O_2, ...O_n]$ for n frames.

Once these features are extracted, we train different classifiers and compare their performance. The classifiers compared are:

- ◇ Probabilistic: Naive Bayes
- ◇ Tree based: J48
- ◇ Ensemble: Random Forest
- ◇ Non-probabilistic: Sequential Minimal Optimization

4.2.1.1 Naive Bayes

Naive Bayes classifier belongs to the family of probabilistic classifiers. The underlying problem of the classifier is to predict a class $c_i \quad \forall i \in [1,k]$ from a set of k classes given a a set of predictors or feature vector, $x = \{x_1, x_2, ...x_n\}$. Naive Bayes achieves this by using the Bayes theorem. The posterior probability, which is the probability of the predictors belonging to a class c_i, is denoted as $p(c_i|x)$ and is calculated as

$$Posterior = \frac{Likelihood * Prior}{Evidence}$$

$$p(c_i|x) = \frac{(\prod_{k=1}^{n} p(x_k|c_i)) * p(c_k)}{\sum_i p(c_i)p(x|c_i)}$$

Finally, the class assigned/predicted class is the one with maximum posterior probability. We use the Naive Bayes algorithm implemented in Weka. Multiple variations of Naive Bayes exist that use different estimation functions for the likelihood. A Gaussian estimator assumes underlying data to belong to a Gaussian distribution. Similarly, there are normal and multinomial estimators. We use the *DiscreteEstimator* that uses data points as is without any assumptions on its underlying distribution.

Naive Bayes is frequently studied and experimented with. However, due to the assumption of strong independence between the predictive variables Naive Bayes is known to be a good classifier but a bad predictor. Each predictor contributes independently to the posterior probability. However, one feature might be dependent on another hence amplifying predictors' contributions. During training, this amplification averages out over the elements in the training set and boosts the classification accuracy. However, during prediction, this effect is prominent as a result the classifier produces poor test instance-class label association.

4.2.1.2 J48 Decision Tree

The J48 decision tree in WEKA is the implementation of the Iterative Dichotomiser 3 algorithm. This algorithm iterates through the attributes present in the dataset and calculates the entropy or information gain for each attribute. The attribute with the maximum information gain is chosen to then split the data and is made a node of the decision tree. The process is repeated for all subsets of attributes. The algorithm requires all of the dataset before it can start building the tree. We use WEKA's implementation of J48 with the default hyperparameters.

4.2.1.3 Hoeffding Tree

Hoeffding tree is an incremental decision tree algorithm. It is advantageous for large data sets because it does not need to see the complete data before making decisions about the branch splits unlike the traditional decision tree algorithm. Moreover, the results produced are statistically comparable to that of a decision tree. It does so by the application of Hoeffding bounds. Hoeffding bounds are concentration inequalities that quantify the convergence given N samples of a population.

Theorem: Let $X_1, X_2, ..., X_n$ be independent bounded random variables. Let $S_n = \sum_{i=1}^{n} X_i$ Then for any $t > 0$ the two sided Hoeffding inequality is

$$P(|S_n - E[S_n]| >= t) <= 2e^{-2nt^2}$$

Let $\alpha = P(|S_n - E[S_n]| >= t)$. α is the probability of making an error for a confidence interval of size $2t$ around $E[S_n]$. Solving $\alpha <= 2e^{-2nt^2}$ for n gives

$$n >= \frac{\log(\frac{2}{\alpha})}{2t^2}$$

Hence, with altleast $\frac{\log(\frac{2}{\alpha})}{2t^2}$ samples, $(1-\alpha)\%$ confidence of $S_n \in [E_{S_n} - t, E_{S_n} + t]$ is guaranteed.

We use Weka's implementation of the classifier with default hyperparamenters.

4.2.1.4 Random Forest

Random forest is an ensemble classifier that uses multiple decision trees built on sub samples of the data set. The advantage of this is that the overfitting by decision trees is controlled and overall accuracy is improved by averaging. We use Weka's implementation of random forest classifier with a default bag size equal to training size. Bootstrap aggregating or bagging weights each model in an ensemble equally. Random forest uses bagging to combine the different decision trees built. In addition it implements *feature bagging* and trains the decision trees on different subsets of the features too. We do not limit the depth of the tree. This works well in case of random forests due to averaging.

4.2.1.5 Sequential Minimal Optimization

Sequential Minimal Optimization (SMO) is an algorithm that is implemented in WEKA to solve the quadratic optimization problem of the Support Vector Machine (SVM) classifier. SVM finds the maximum margin hyperplane that divides a dataset of n points into two classes. The QP of SVM is

$$\begin{aligned} & Minimize \quad \|\vec{w}\| \\ s.t. \quad & y_i(\vec{w}\cdot\vec{x}_i - b) \geq 1, \qquad \forall i = 1,\ldots,n \end{aligned}$$

The hyperplane is given by $\vec{w}\cdot\vec{x}_i - b$ where $\vec{w}$ is normal to the hyperplane and $\frac{b}{\|\vec{w}\|}$ is the offset of the hyperplane from the origin.

SMO is an algorithm that breaks the above optimization problem to smaller ones and iteratively solves each using the Lagrange multipliers. Multi-class problems are solved pairwise using the one-vs-one strategy. Hence, for a c class classification, $\frac{c.(c+1)}{2}$ binary classifiers are created and the class with maximum number of 1's predicted is assigned to the test data.

4.2.2 Time-series Similarity

In section 4.2.1, structural features were extracted from the time series and used to train classifiers. In this section, we use the raw CEUS brightness values that build the time series. The CEUS brightness values of a tissue class over n frames is collected. For an instance we use the most recent 20 frame values for every class of tissues as numeric attributes. Additionally, one nominal attributing denoting tissue class is added. We do not include the first 25 frames as the CEUS images do not show up until much later in the videos. For training, one instance per class per frame is created for all $(n-25)$ frames of the video. For testing, the time-series is made pixel-wise - an instance is created per pixel in the frame with attributes being the CEUS brightness values of the pixels seen in the last 20 frames.

The generated time-series are compared with each other using the dynamic time warping (DTW) distance between them. DTW is a time-series alignment algorithm that computes the minimum distance between two series. It differs from traditional distance measures like the Eucledian distance in that it is one to many and not one to one. When calculating the Eucledian distance between two series, the distance between two aligned points (one-to-one) is calculated and summed up. On the other hand, in DTW, two time series are aligned and the distance between one data point on a series to all points on another series is found resulting in a distance matrix. Figure 4.5 illustrates the distance matrix between two series $A = a_1, a_2, ..a_n$ and $B = b_1, b_2, ..., b_m$. Both series start on the bottom left of the matrix and end on the top right. A path between these two points is created by selecting the minimum distance at each aligned data point. The DTW between the two series is the sum of distances along this path.

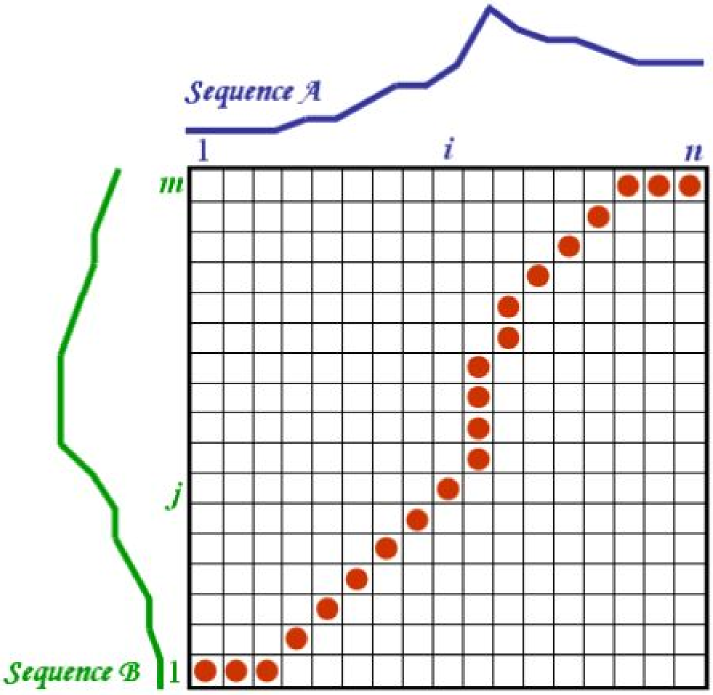


Figure 4.5: The distance matrix of aligned time series A(t) and B(t) with the minimum distance path shown in red. [8]

A k-Nearest Neighbor (k-NN) classifier is learnt based on the DTW distance metric. The Ibk class provided by WEKA implements the k-NN algorithm. K-NN is a simple algorithm that uses the neighborhood of a test data-point for classification or regression. In the case of classification, the class assigned to the data point is the majority of its neighbor's classes. For regression, the value predicted is given by the average of the neighbors. LinearNNSearch is the default search algorithm that WEKA uses to find the neighbors. A custom distance function can be set for the algorithm. However, there was no available implementation of a DTW distance function class for WEKA. We created one using inspiration from the Java implementation of the algorithm in [9].

4.3 Deep Learning Approach

Dietrich et al. [32] list the various factors that the CEUS output depends on. These can be broadly divided into equipment based factors and patient based factors. Inclusion of such a broad spectrum of variablity as human extracted features for tissue recognition is time-consuming and highly succeptible to being errorneous. Further, some factors of variablity differentiate tissues better than others. Hence, feature selection also becomes important. Deep learning provides a solution with two-fold benefits. It not only identifies underlying features but also chooses the most differentiating ones for learning. Hence, we also implement a deep learning based pipeine for tissue recognition.

4.3.1 Brief overview

Neural networks consist of many connected processors called neurons, each producing a sequence of real-valued activations. Figure 4.6 shows a neuron, in the simple perceptron neural network configuration, with n inputs $x_1, x_2, x_3, ...x_n$ having corresponding weights $w_1, w_2, w_3, ...w_n$. The contribution of each input by weight is aggregated. Optionally a *bias* can also be added. The result is passed to the neuron's activation function f. The activation function can be any function like the *linear, Rectifier(ReLU), tanh, sigmoid or softmax*. Linear functions are used only when the data for classification is known to be linearly separable. As this is not usually the case, in most cases non-linear activation functions are adopted. Deep Learning networks consist of multiple layers of neurons. The goal of supervised deep learning is to learn the weights and biases given an input and the desired output.

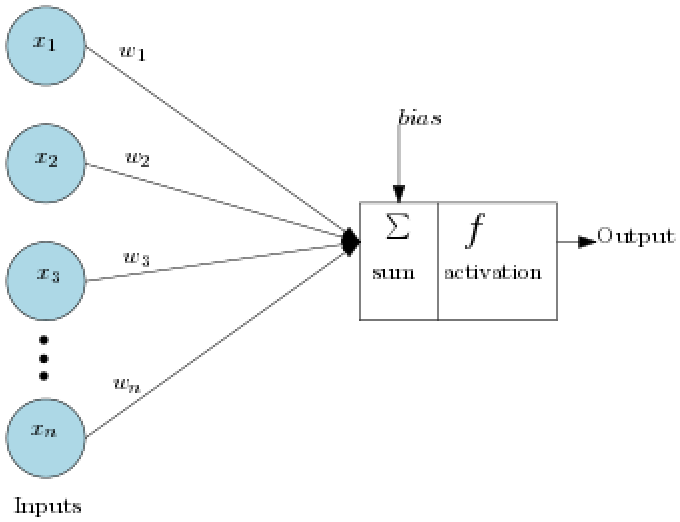


Figure 4.6: An artificial neuron in a Perceptron neural network.

We implement a deep network in python using Keras (https://keras.io/) with a Tensorflow backend. Keras is a high-level neural network library that is flexible, modular and extensible. It supports multiple low-level deep learning libraries closer to the machine like Tensorflow by Google, CNTK by Microsoft and Theanos and provides APIs to interact with the backend libraries too. Although it supports various backends, the library recommends using TensorFlow. Hence, we use it with a TensorFlow backend.

4.3.2 Data Augmentation

We were provided with a couple of annotated frames per video. Since the position of tissues changes per frame, one annotation file could not be used for all frames in a video. Each frame needed to be annotated independently. A single patient's video consisted of upto 409 frames. Labelling every frame for multiple patients was extensively time-consuming. Hence, we leveraged data augmentation.

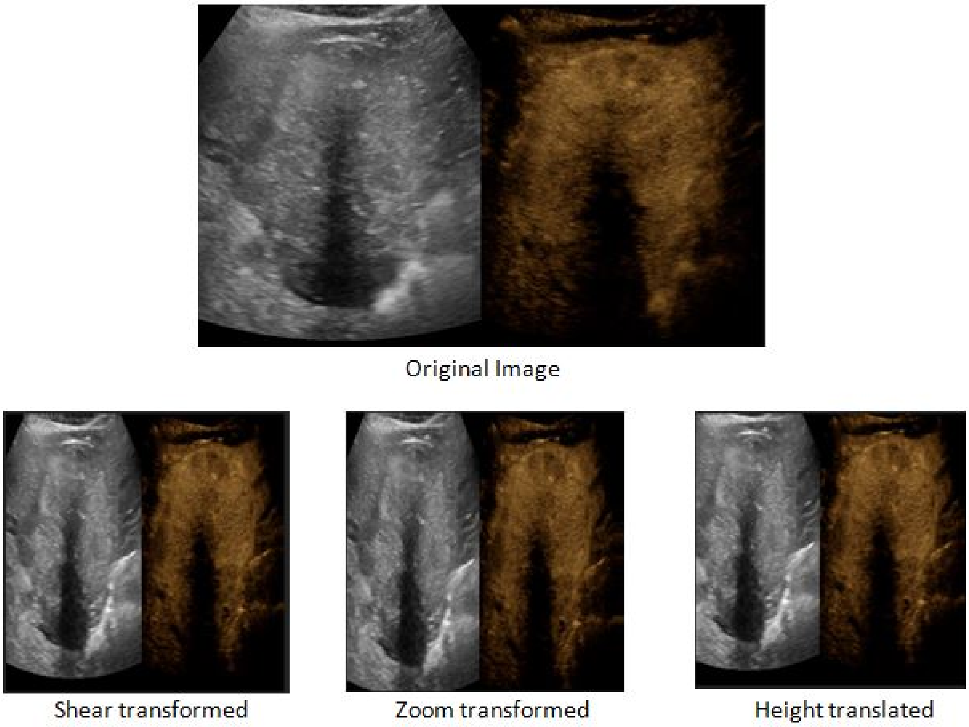


Figure 4.7: Augmented images.

We chose 4 patients with relatively short sonograph videos having varying tumor tissues and at different stages of HCC. These amounted to a total of 1046 frames. Further, from the evenly distributed frames across each of the videos, few frames were picked as representatives of the video. Since the frames picked were from different time points in the video, a sample for each phase of the contrast enhancement was considered. We capped the maximum frame number picked from any video at the frame count of the shortest video (163). Hence, the representative frames were picked from only the first 163 frames of every video. This truncation is needed to ensure the scale of the ”time” factor across the videos is the same.

This selection of representatives resulted in a total of 193 frames from the 4 patients. These frames were annotated by hand using the Bbox Labelling tool described in section 4.2. However, we labelled the images with only two classes in this approach- tumor and non-tumor.

It is widely accepted that a large training set is a pre-requisite for a well-performing neural network model. When the training size is small, the model tends to overfit. In order to increase the number of training samples and include invariance and robustness to the model, we applied translational and rotational variance to the training images. These are chosen keeping in mind the shaking and real-time nature of ultrasound. Keras provides dedicated functions for image preprocessing. We use the ImageDataGenerator class to perform the augmentation of training images. We apply both height and width shift translation on the training images. Parameters ”*height_shift_range*” and ”*width_shift_range*” are both set to 0.05. Keras picks a random number, *x*, from the uniform distribution in the interval [-0.05, 0.05] and translates the image pixels by *x* along the image height or width. A rotational range of 0.2 is used and a zoom and shear range of 0.05 each is applied. The augmentations are applied in batches of 3 images to the training set. We leverage the ”generator” model of Keras where the data is not loaded all at once but instead in batches. Figure 4.7 shows some augmentations of a training sample.

4.3.3 U-Net

Convolution Neural Networks (CNN) are widely preferred for tasks involving image processing. They have been successful is identifying features not readily visible to the human eye. U-net, which is a deep convolution network, built by Ronneberger et al. [65] in 2015 outperforms traditional CNN networks and has since been extensively used for medical image processing. Our implementation choices are discussed and described in the following sections. The graph of the U-Net implemented is shown in figure 4.13 (a)-(c) (in alphabetical order).

4.3.3.1 Input Layer

The input layer is the starting point of the architecture. Every frame of the sonograph video is preprocessed before it is fed into this input layer. A frame is converted into a 3 dimensional numpy array. The first two dimensions carry the spatial information i.e. the x-y coordinates of the pixels present and the third dimension denotes the ’depth’ or the number of color channels in an RGB image. In our case, this is 3. This can be visualized as a 2-d numpy array for each color channel stacked on top of one another.

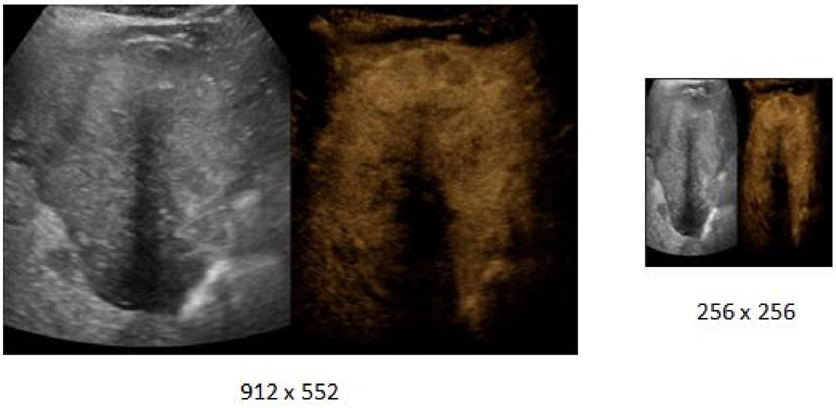


Figure 4.8: A sonograph frame resized to be fed as input to the U-Net.

The input image size can affect the training time and the performance of the network. In order to strike a good balance between both, we chose to resize our images to 256 x 256 from the original 912 x 552. The reduction in size, benefits the train time. Also, since the model learns several features to describe the images, it is generally assumed it takes care of learning features associated with size and aspect. Several deep learning libraries such as TensorFlow also adopt image resizing before training. The resizing was implemented using the *ImageDataGenerator* class provided by Keras. Hence, the dimension of the input tensor is $(256, 256, 3)$. An example of an input frame is shown in figure 4.8.

4.3.3.2 Convolution Layers

We use the Conv2D layer provided by Keras to implement the convolution layers. We use the standard 3 x 3 convolution filters with a stride of 1. In order to prevent loss of information during the convolutions k-zeros padding technique is used. At all times, the image size is maintained at 256 x 256. Hence, k is calculated appropriately. Therefore, we set parameter padding as ’same’ to ensure the the input is padded before the filter is convoluted on it.

Once, the filter is applied of the input image, the resulting tensor is passed through an activation function. The activation function used is Rectifier Linear Unit (ReLU). This function preserves only the positive parts of an input. That is, $f(x) = x^+$. Initializers are needed to define the initial weights of the neurons in a layer. We use the *He normal* initializer that randomly picks weights from a normal distribution with mean 0 and standard deviation $\sqrt{\frac{2}{in}}$ where *in* is the number of input units to the layer.

4.3.3.3 Max Pooling Layers

For pooling the usual window size used is (2 x 2) and we stick to this convention. The window is applied on the image with a stride of 1. Our implementation has 4 Max pooling layers with window sizes of 2 x 2 defined by MaxPooling2D function provided by Keras. The max pooling operation is applied to each color channel in the data separately. An example of max pooling is shown in figure 4.9.

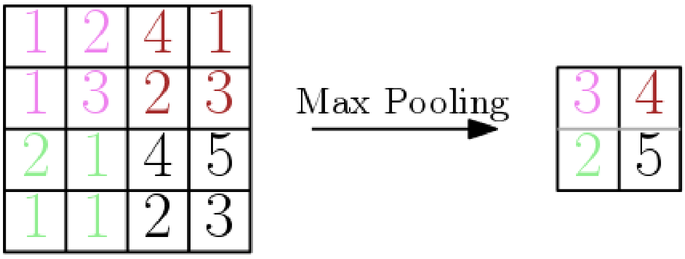


Figure 4.9: Max pooling with window size 2 x 2.

4.3.3.4 Dropout Layers

Dropout layers drop some activations in every layer by setting them to zero. This ensures that the model doesn’t overfit the training data and perform badly on test data by inclusion of redundancy. Srivastava et al. [76], who are the inventors of the dropout technique, carry out all the experiments with a dropout probability of 0.5 for the hidden layer units. The models with dropout perform better on *all* the datasets tested on including MNIST and ImageNet. They state that a dropout rate of 0.5 for the hidden units was often found to be optimal. Hence, we choose to use a dropout rate of 0.5.

4.3.3.5 Up-Convolution Layers

U-Net uses up-convolution layers to improve localization in the images resulting in improved segmentation and learns the ”where” in the data. As is common practice, we use a 2 x 2 filter for the up convolution. Every value of a cell is repeated four times in the feature map generated. This is filter is slid over the input map with a stride of 2. An example of the same is shown in figure 4.10. The up convolution is implemented in Keras using the UpSampling2D layer provided. We use the default *nearest* interpolation.

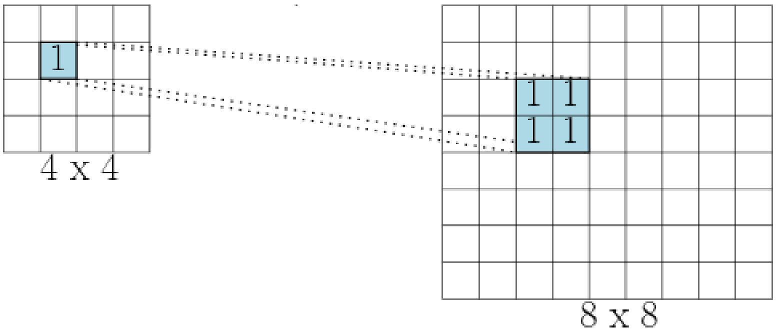


Figure 4.10: Up-convolution with a window size of 2 x 2.

4.3.3.6 Output Layer

The expected output implemented network is a segmented image marked with the tissue class. Figure 4.11 shows the input to and expected output from the implemented deep network. While testing, the segmented frames derived from the model are overlaid on the test frames and presented as the final result.

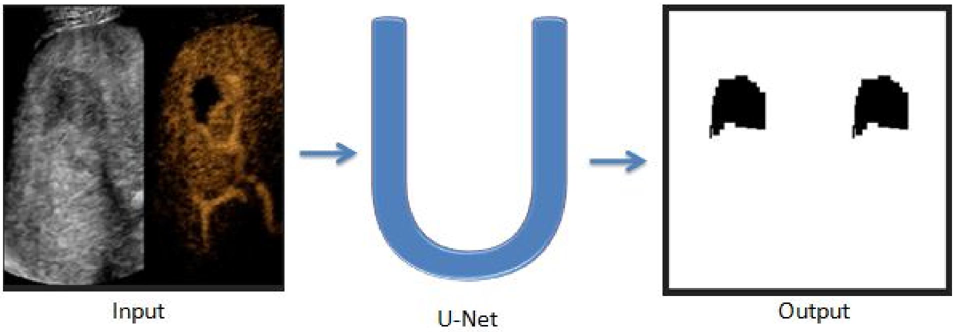


Figure 4.11: Input and outputs used to train the Unet.

We use a 1 x 1 x 3 convolution to reduce the number of feature channels in the penultimate layer of the network to 3. Although the number of classes in the segmentation are two - tumor and non-tumor, we maintain the 3 RGB channels to ease the next overlaying step.

4.3.3.7 Loss Function

Loss functions are chosen based on the underlying operation of the neural network - regression or classification. Because classification tasks produce a probabilistic outcome loss functions like binary cross-entropy or negative log-likelihood are preferred over others. Considering one-hot encoding, the target variable in classification tasks are binary - 0 or 1 for "present in class" and "not present in class" respectively. In our case, the number of classes are 2. Also, the two classes are independent. Hence, we use the binary cross-entropy loss function. It provides the advantage of fast conversion and is more likely to reach global optimization. Cross-entropy tends to allow errors to change weights their derivatives during back-propagation are asymptotically close to 0 [60].

4.3.3.8 Optimization Function

Adaptive learning algorithms use heuristics and adapt the learning rate and do not require them to be manually tuned. Adaptive Moment Estimation (ADAM) is an optimization algorithm that adapts learning rate scale for different layers and is found to be cost-effective and fast.

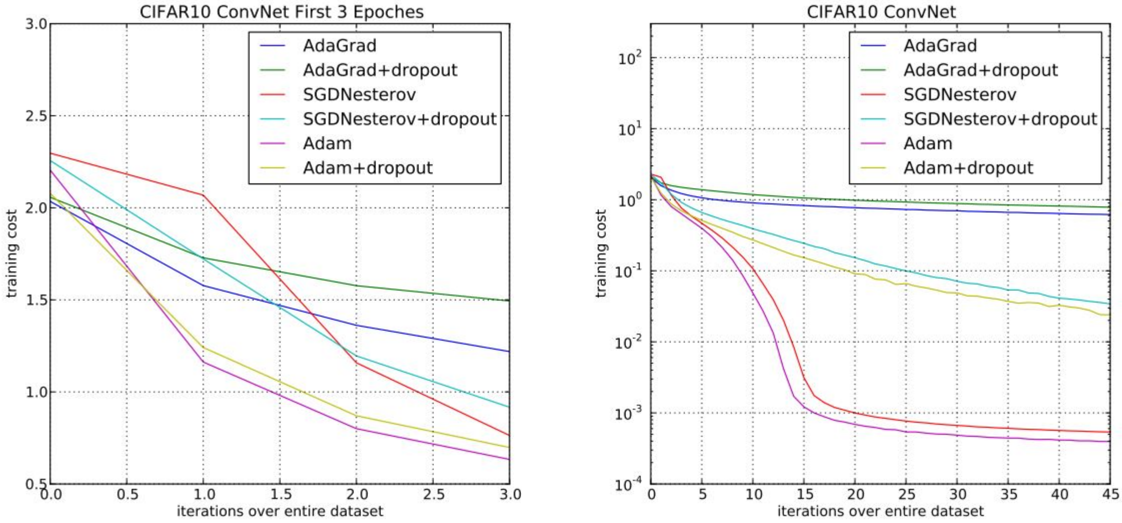


Figure 4.12: Training cost of various optimization algorithms over a CNN with 3 alternating 5 x 5 convoluntion and 3 x 3 max-pooling layers applied on the CIFAR-10 image dataset [46].

As seen in figure 4.12, both during the initial epochs of the training and later, ADAM and SGD converge but with the ADAM having marginally lower training cost. Hence, we chose ADAM as the optimization function.

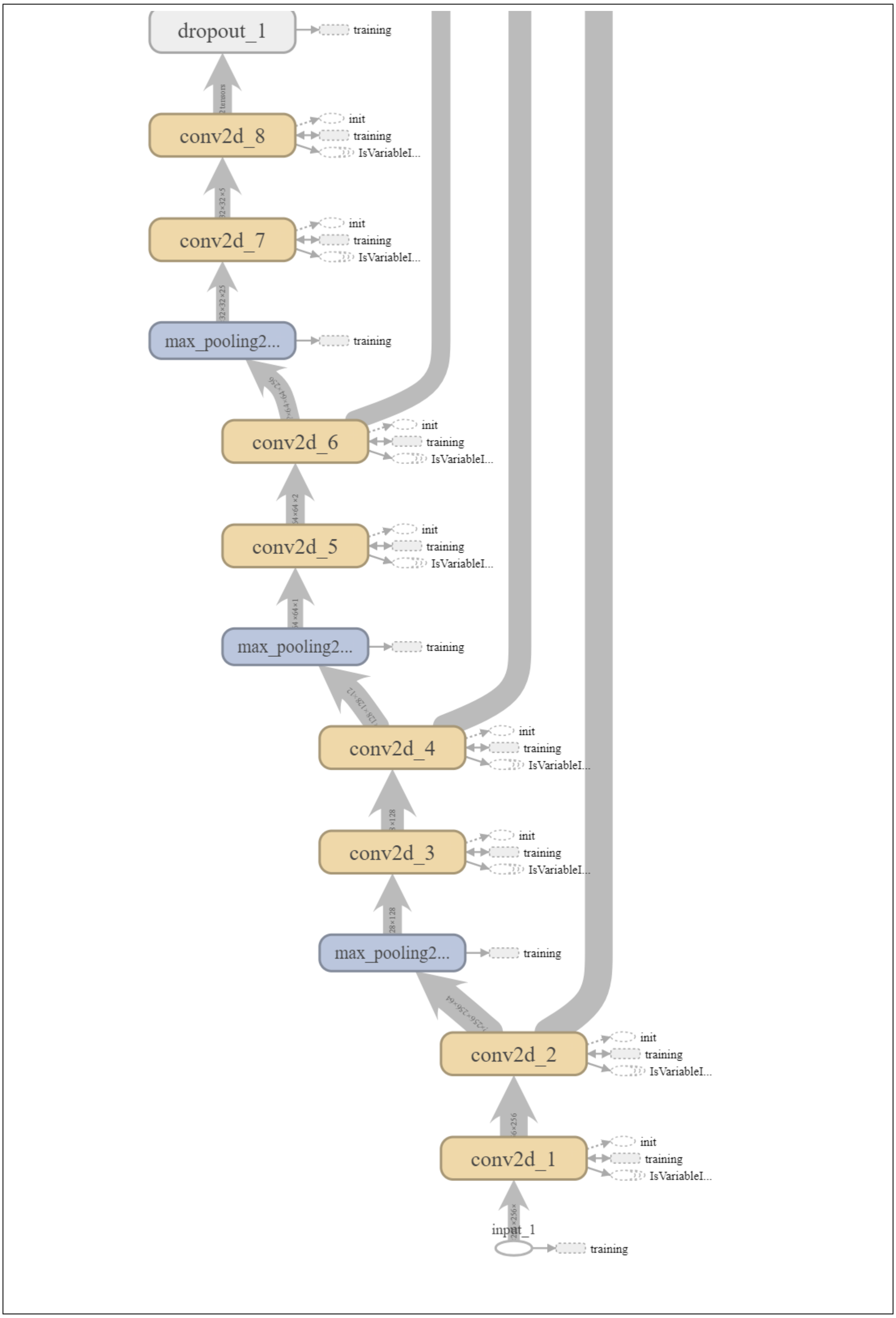
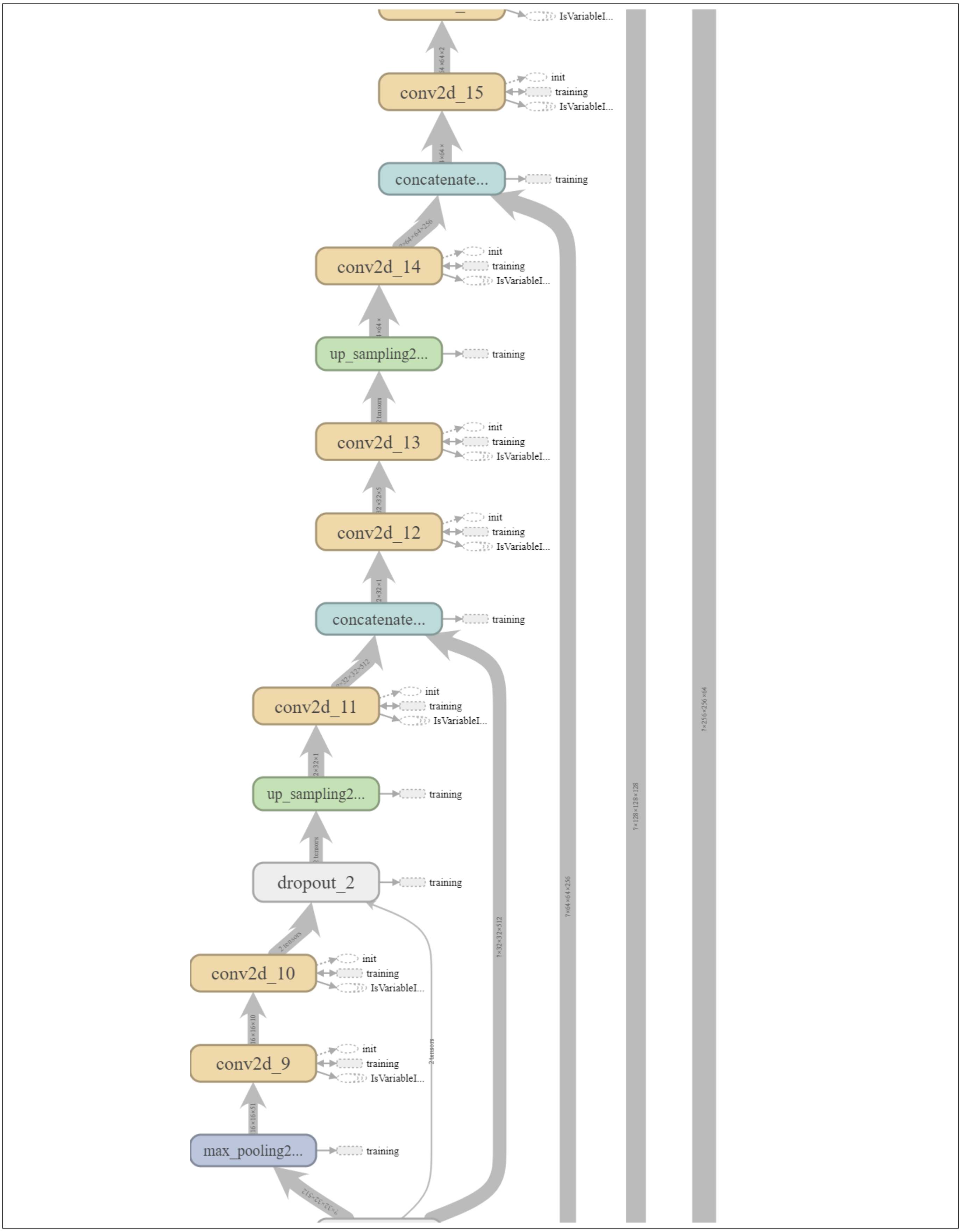
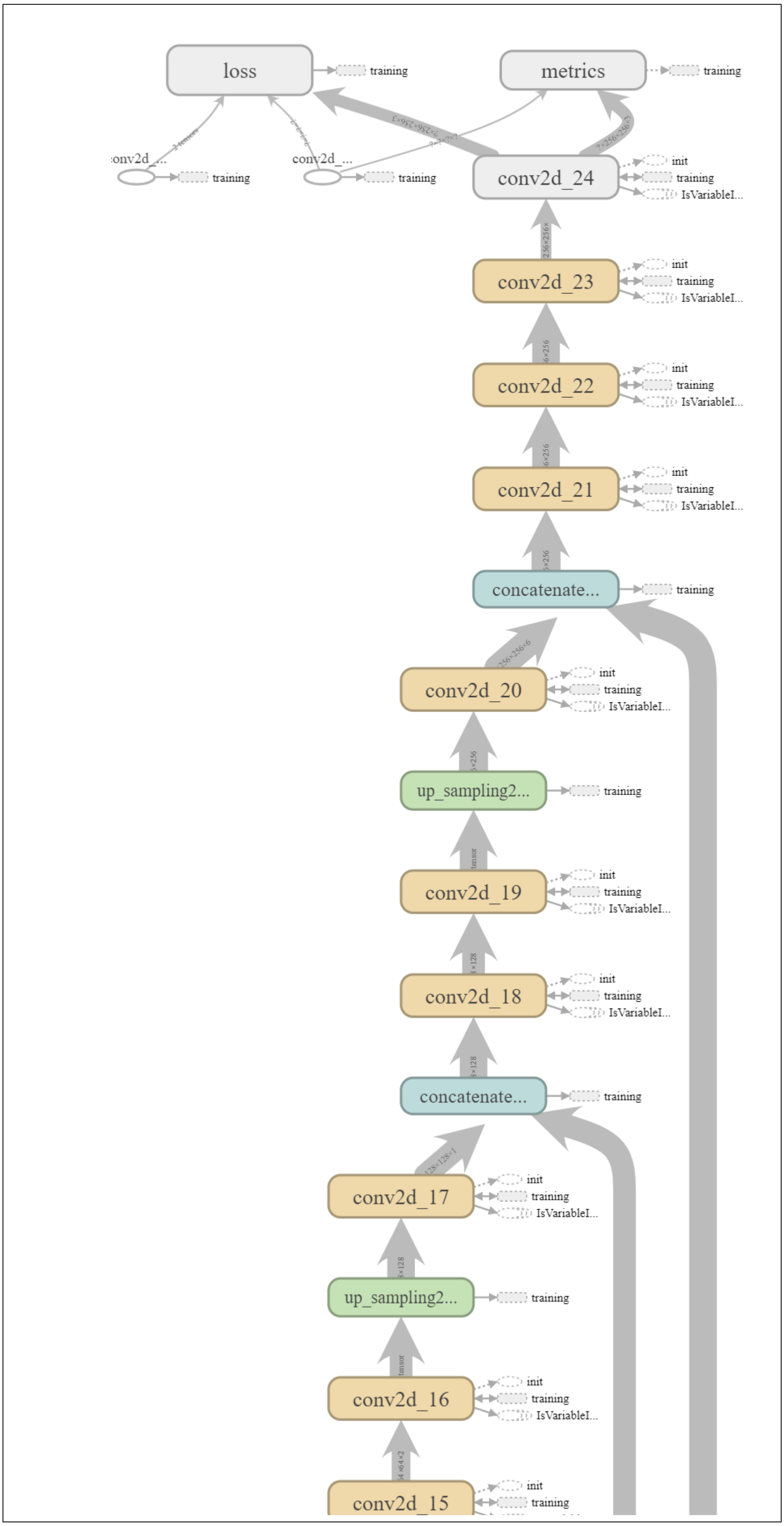


Figure 4.13: (a) U-Net implemented - The input layer begins.



: Figure 4.16 (b) U-Net implemented - Contraction convolution layers.



: Figure 4.16 (c) U-Net implemented- Expansion convolution layers.

5

Evaluation

The experiments conducted and an evaluation of the obtained results are presented in this chapter. First, the evaluation metrics used are explained briefly in section 5.1. Later, the two sections - 5.2 and 5.3 explain the evaluation set-up and results of the two main approaches - traditional feature extraction based machine learning and deep learning.

5.1 Evaluation Metrics

There are multiple well-known techniques to evaluate the performance of machine learning models. The confusion matrix summarizes the model's performance. In the scenario of binary classification, the confusion matrix delivers the information about true positives and true negatives - test instances belonging to positive and negative classes, respectively, and that were correctly predicted; false positives - test instances that belonging to the positive class and were wrongly predicted to belong to the negative class and false negatives - test instances belonging to the negative class and were incorrectly predicted to belong to the positive class. In most cases, especially in the medical, monitoring false positives is very critical in evaluating a model ([73], [83] and [23]).

Precision accounts for the correct predictions made. It is computed by true positives(TP) divided by the sum of true positives and false positives(FP).

$$Precision = \frac{TP}{TP+FP}$$

Recall or sensitivity denotes the number of ground truth classes that were accurately predicted. It calculated by true positives (TP) divided by the sum of true positives and false negatives (FN).

$$Sensitivity = \frac{TP}{TP+FN}$$

F1 score is a measure that combines both precision and recall. It is the harmonic mean of precision and recall.

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

The accuracy of a classifier is the fraction of test instances correctly classified.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

The Receiver Operating Characteristic (ROC) curve is a plot between the true positive rate (TPR or recall) and the false positive rate (FPR). The range of each axis is $[0,1]$. The closer the area under the ROC (AUC or AUROC) is to 1, the better the classifier. In figure 5.1, classifier with ROC marked C1 has a close to linear relationship between TPR and FPR. Hence, what is does is very close to guessing. Whereas, classifier $C2$ clearly has some instances where the FPR beats TPR and vice-versa. Therefore, it is able to distinguish classes with confidence. Also, the area under the curve for $C2$ is greater than $C1$.

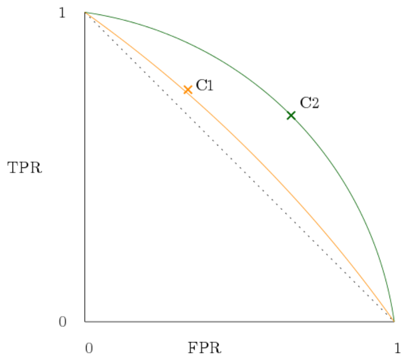


Figure 5.1: Based on the AUC, classifier C2 performs better than classifier C1.

These evaluation metrics are based on the fact that the ground truth is known accurately. This implies the importance of correct labelling in supervised learning. Therefore, the collected data is labelled strictly according to the information from experts from the University Hospital Münster (UKM).

5.1.1 Intersection over Union

Intersection over union (IOU) is an important evaluation measure used for image segmentation. It is the ratio between the intersection of actual and predicted image regions and their union.

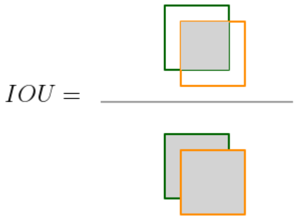


Figure 5.2: Intersection over union metric.

5.2 Time-series Approach

5.2.1 Evaluation Setup

For the time series approach, the sonograph videos of two patients were used. The two patients had tumors at different stages. Figure 5.3 shows frames from the songraphy of the two patients.

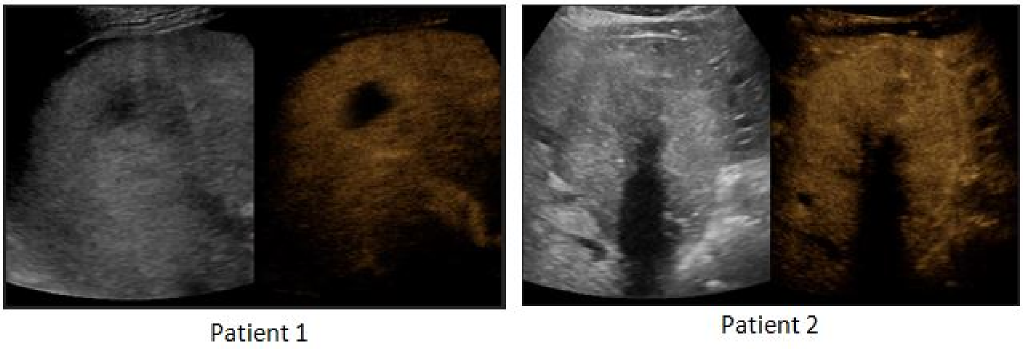


Figure 5.3: Tumors of two patients at different stages of HCC.

We received information about the tissues present in the videos from the physicians. The information received is shown in figure 5.4.

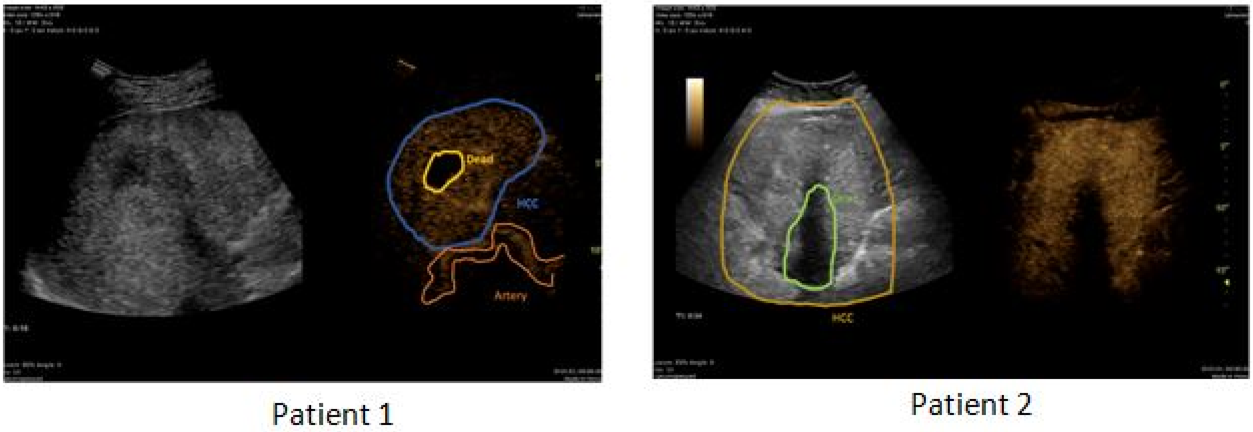
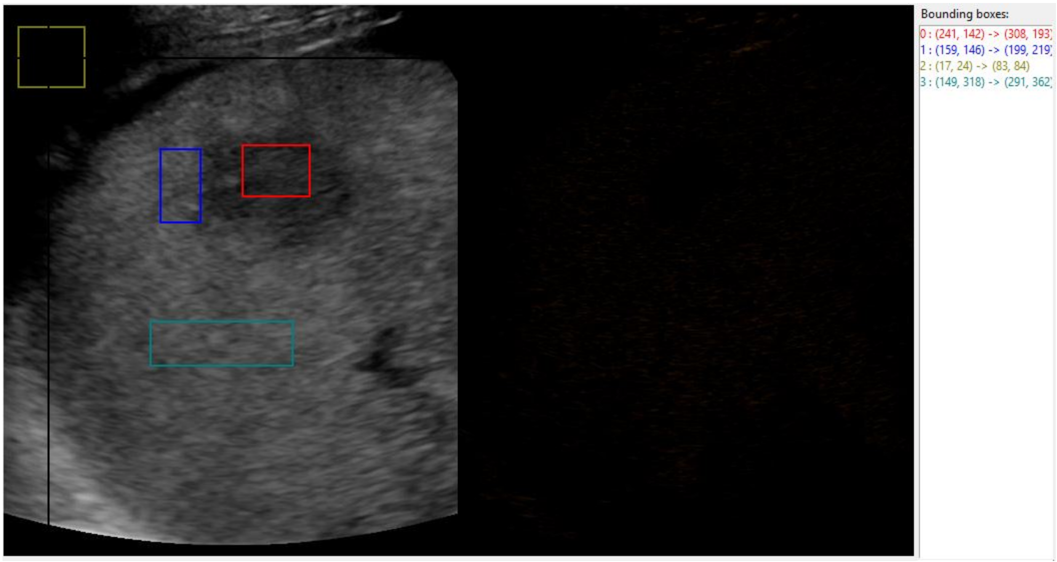


Figure 5.4: Tissues marked as per information from physicians.

Hence, we labelled the frames with four classes. Figure 5.5 shows the classes for a frame of patient 1. This is a rather simple example. In majority of the frames, multiple rectangles had to be used to mark a class.

- ◇ Class 0- This class denotes the dead tissue. It corresponds to the highly affected area
- ◇ Class 1- Also part of the tumor tissue.
- ◇ Class 2- Area not part of the sonograph. Pixels that are in the black bounding regions.
- ◇ Class 3- This class marks the normal tissue.



Bounding boxes:

0 : (241, 142) -> (308, 193

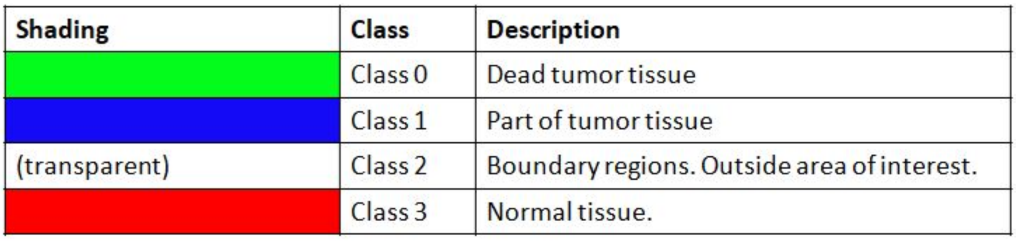
1 : (159, 146) -> (199, 219

2 : (17, 24) -> (83, 84)

3 : (149, 318) -> (291, 362

Figure 5.5: Example of the four classes marked on a sonograph frame. The class label and the rectangle coordinates are listed on the right. Each color signified a class.

In the classification results we use the color encoding shown in figure 5.6 to mark the classes identified.



Shading	Class	Description
🟩	Class 0	Dead tumor tissue
🟦	Class 1	Part of tumor tissue
(transparent)	Class 2	Boundary regions. Outside area of interest.
🟥	Class 3	Normal tissue.

Figure 5.6: The color encoding used when classifying the frames.

Labelled ultrasound videos of two patients were used to train different models using the time series features extracted. The features that were extracted are listed in table 4.2 in section 4.2.1. We first built idiosyncratic models for each patient and then combined the videos to build a single model for both patients. The results of the various algorithms applied are presented in this section.

5.2.1.1 Combining Videos

Combining patient videos of different sizes together can increase the bias in the data. The model can be biased towards the videos with more number of frames. As of the above chosen patients, patient 1's sonograph has 404 frames while that of patient 2 has only 202 frames. One method to not introduce unnecessary bias is truncation. We adopt this technique in the current study. Notable advantages of truncation are ease of implementation. We simply look at only as many frames as present in the shortest video. This makes it a viable option when combining a large number of videos. However, truncation comes with its drawbacks- loss of information. In scenarios where the sonographs are long, the initial frames usually have no CEUS information.

Hence, truncating from the beginning of videos is not a very good idea. To overcome this, a section of the video with highest CEUS information can be chosen. Although this introduces human bias when ”selecting” the most important section, it prevents using frames carrying low information. Another method that can be used is interpolation. Although it prevents loss of information, the method can be cumbersome when combining multiple videos. Also, new data is highly dependent on the goodness of the interpolation function.

As in our current experiment, we are looking at only two patients, we chose the truncation method and truncate the video of patient 1 to 202 frames starting from the first. Hence, the combined video consists of 404 frames. As we look at 4 representative pixels per frame; one corresponding to each class, the total number of training instances from the combined data are 1616. As for patient 2 alone, we train with 808 instances (202 frames * 4 classes). On the other hand, models with trained only with data from patient 1 work with 1616 instances (404 frames * 4 classes).

5.2.1.2 Single Patient Models

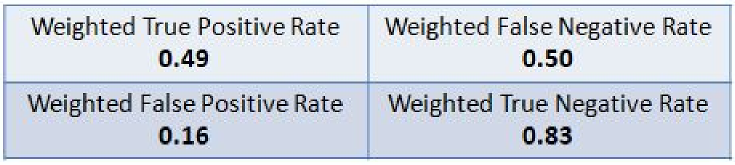
The models that are trained on single patients can be mistakenly thought to not be ”generalized”. However, it is to be noted that the models are trained with pixels and not frames. The training ”pixels” are only 4 per frame- one corresponding to every class of tissue described earlier. While testing, all the pixels of every frame are looked at. Hence, the models for single patients are certainly not meagre ”memorization”.

5.2.2 Evaluation Results

5.2.2.1 Naive Bayes

Combined

We performed 10 fold cross validation for the model to evaluate its goodness and robustness to changing training data. The false positive rate is an important measure as misclassifying a non tumor tissue as tumor is highly undesirable. The weighted false positive rate is 0.16. The weighted prediction rates observed are summarized in table 5.1

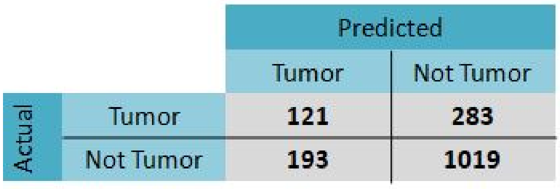


Weighted True Positive Rate **0.49**	Weighted False Negative Rate **0.50**
Weighted False Positive Rate **0.16**	Weighted True Negative Rate **0.83**

Table 5.1: Classification rates for the Naive Bayes classifier on combined data from 2 patients.

We also look at the confusion matrix for class 0 which we denote the tumor tissue by as a measure of goodness of the classifier. The confusion matrix for the class 0 generated during the 10 fold cross validation is shown in table 5.2. The false positive rate is low at 0.15. However, the true positive rate also drops to 0.29. The effect of this is reflected in the predicted images in figures 5.10 and 5.11 which fail to identify the tumor tissues.

Further, the kappa static which is a better measure than accuracy is 0.32. It is found that 50.3% of the cases are incorrectly predicted with a low weighted precision and weighted recall



		Predicted	
		Tumor	Not Tumor
Actual	Tumor	121	283
	Not Tumor	193	1019

Table 5.2: Confusion matrix for tumor tissue labelled class 0.

of 0.55 and 0.49 respectively. Moreover the weighted AUC was observed to be 0.75 while that of tumor class 0 was 0.62 as shown in figure 5.7.

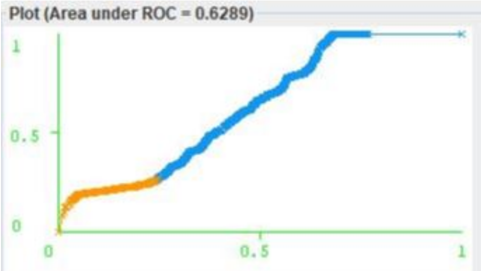
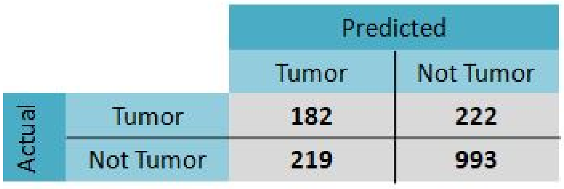


Figure 5.7: Area under ROC for Naive Bayes model on the combined data of two patients for tumor class (class 0).

Patient 1

The Naive Bayes classifier trained with time series features of the CEUS brightness in the sonograph video of patient 1 has a kappa statistic of 0.34. It predicts incorrectly 49.3% of times. The weighted precision and weighted recall are 0.51 and 0.50 respectively. On the other hand, the weighted false positive rate is seen to be 0.16 like the combined model. The weighted AUC is 0.84 which is much higher than the combined videos' Naive Bayes classifier. The confusion matrix for tumor tissue is shown in figure 5.8. The false positive rate for the tumor class is stands at 0.18 which is a little high in comparison to the combined model.

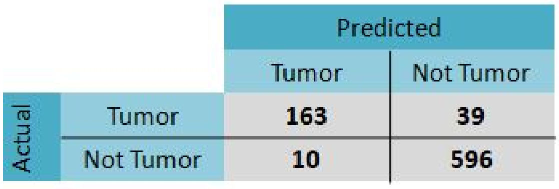


		Predicted	
		Tumor	Not Tumor
Actual	Tumor	182	222
	Not Tumor	219	993

Figure 5.8: Confusion matrix for the tumor class in Naive Bayes classifier built with patient 1's data.

Patient 2

The classifier trained with the CEUS video of patient 2, shows a hight correct prediction rate percentage of 68.19% and kappa statistic of 0.58. The weighted false positive rate drops to 0.11 and the weighted AUC is 0.83. The confusion matrix for class tumor is shown in figure 5.9. We calculate that the false positive rate for this class is 0.02.



		Predicted	
		Tumor	Not Tumor
Actual	Tumor	**163**	**39**
	Not Tumor	**10**	**596**

Figure 5.9: Confusion matrix for the tumor class in Naive Bayes classifier built with patient 2's data.

Summary

In the model trained on the combined data from the two patients, the performance of the model on both patients is found to be rather poor. When tested with patient 1's data, the model identifies tissues belonging to class 3 (representing normal tissues, encoded in red) correctly as shown in 5.10. However, the other results of the other classes are poor and the results do not improve as the number of frames seen increases.

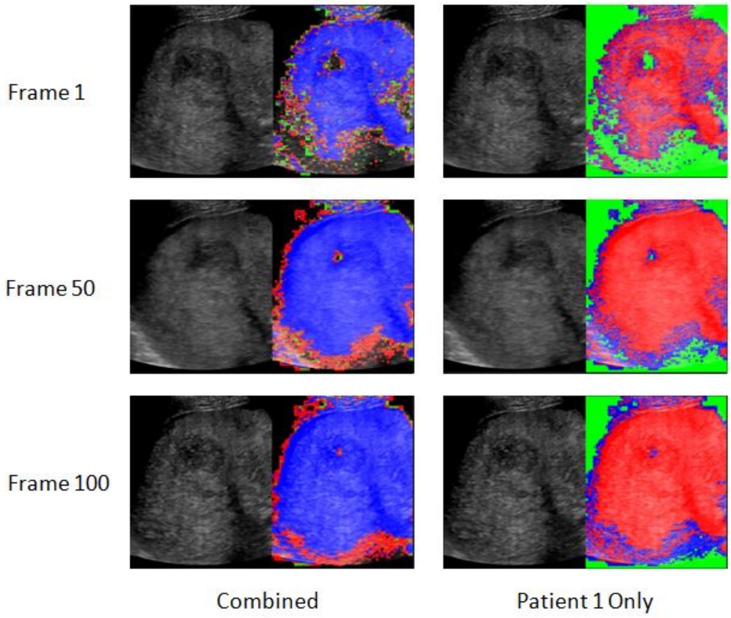


Figure 5.10: Frames of patient 1's CEUS video as segmented using the Naive Bayes model trained with combined data and data only from patient 1.

As for patient 2's sonograph, although the dark parts in the boundary of the image are rightly classified as class 2, the model identifies the majority of the frame as class 1 - part of the tumor. Figure 5.11 shows the same.

Overall, the individual patient model for patient 1 not seem to be much better than the classifier trained on the combined data of two patients both in terms of segmented test frames and in terms of cross validation statistics. The model rained on patient 2's data alone, shows significantly lower false positive rates. The 10FCV statistics are summarised in table 5.3.

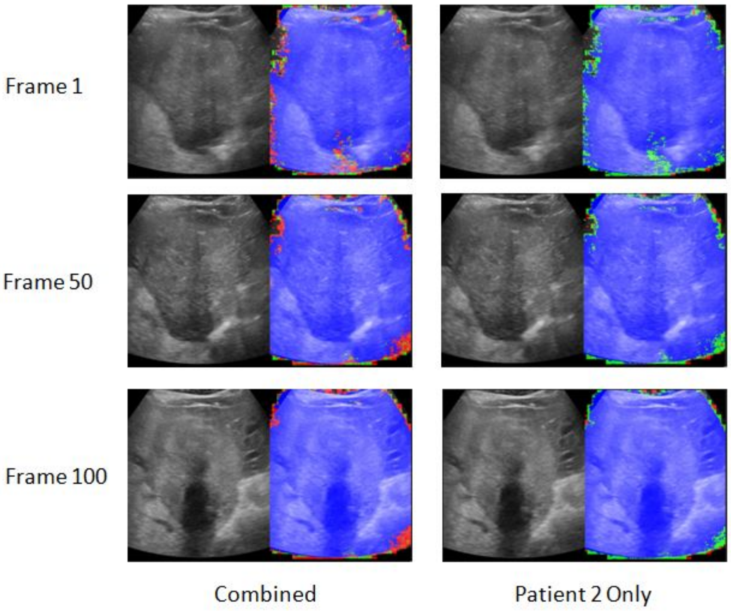


Figure 5.11: Frames of patient 2's CEUS video as segmented using the Naive Bayes model trained with combined data and data only from patient 2.

Data	%Correct	%Incorrect	Kappa	Weighted AUC	WTP	WFN	WFP	WTN	WPrecision	WRecall
Combined	49.69	50.31	0.33	0.75	0.5	0.5	0.17	0.83	0.56	0.5
Patient 1	50.68	49.32	0.34	0.85	0.51	0.49	0.16	0.84	0.51	0.51
Patient 2	68.19	31.81	0.58	0.83	0.68	0.32	0.11	0.89	0.73	0.68

Table 5.3: Evaluation statistics for the Naive Bayes Classifier.

5.2.2.2 J48 Decision Tree

The decision trees are generated for training data from patient 1, patient 2 and both combined. The three cases are discussed below.

Combined

The decision tree classifier trained on the combination of two patients' data is shown in figure 5.12.

The 10FCV results in a weighted AUC of 0.91 and a tumor class AUC of 0.87 as shown in figure 5.13. 79.08% instances are correctly classified by the classifier. The weighted precision and recall values are 0.88 and 0.79 respectively. Owing to which, the kappa statistic is also hight at 0.72. The model also shows a low weighted false positive rate of 0.07. However, the confusion matrix of the tumor class in figure 5.14 shows that the true positives and false positives are quite similar in number. Hence, pointing towards a classifier that just ”guesses” the tumors.

Patient 1

The decision tree generated for patient 1 is visualized in figure 5.16. The differentiating features, providing the most information gain are identified to be ”aboveMean” and ”belowMean”.

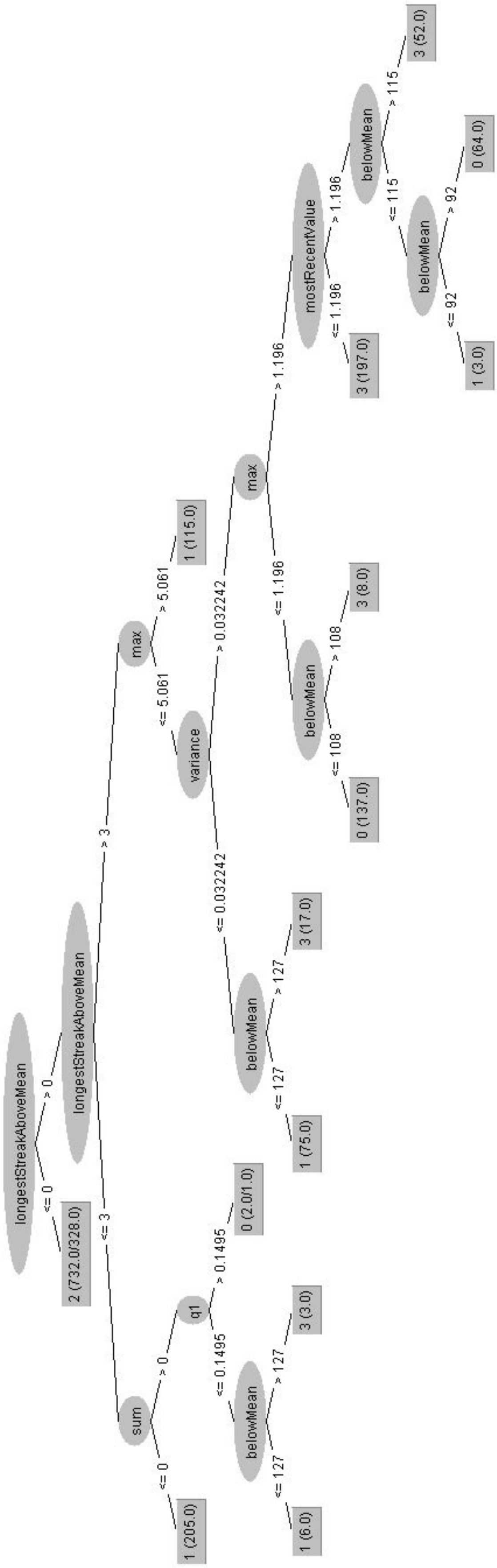


Figure 5.12: The J48 decision tree for combined data of two patients.

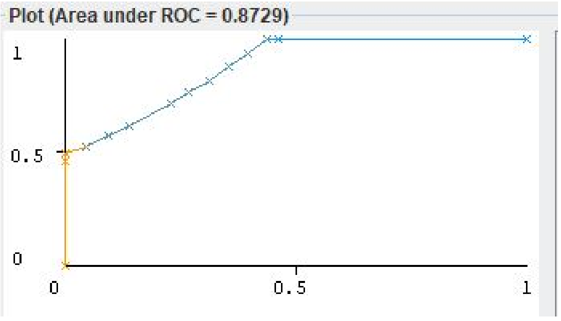
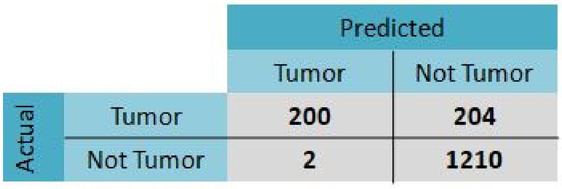


Figure 5.13: AUC of the tumor class (class 0) for combined data using the J48 Decision tree.

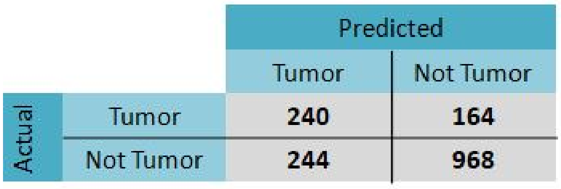


		Predicted	
		Tumor	Not Tumor
Actual	Tumor	200	204
	Not Tumor	2	1210

Figure 5.14: The confusion matrix for the tumor class.

These correspond to the number of instances with CEUS brightness values above and below the mean respectively.

The model classifies 75% instances correctly. It demonstrates weighted precision and recall of 0.75 each. The kappa statistic lowers to 0.66. As for the confusion matrix of the tumor class, the prediction rates are not very different from the combined model. The confusion matrix for the tumor class is shown in figure 5.15.



		Predicted	
		Tumor	Not Tumor
Actual	Tumor	240	164
	Not Tumor	244	968

Figure 5.15: The confusion matrix for the tumor class with the J48 decision tree classifier on patient 1's data.

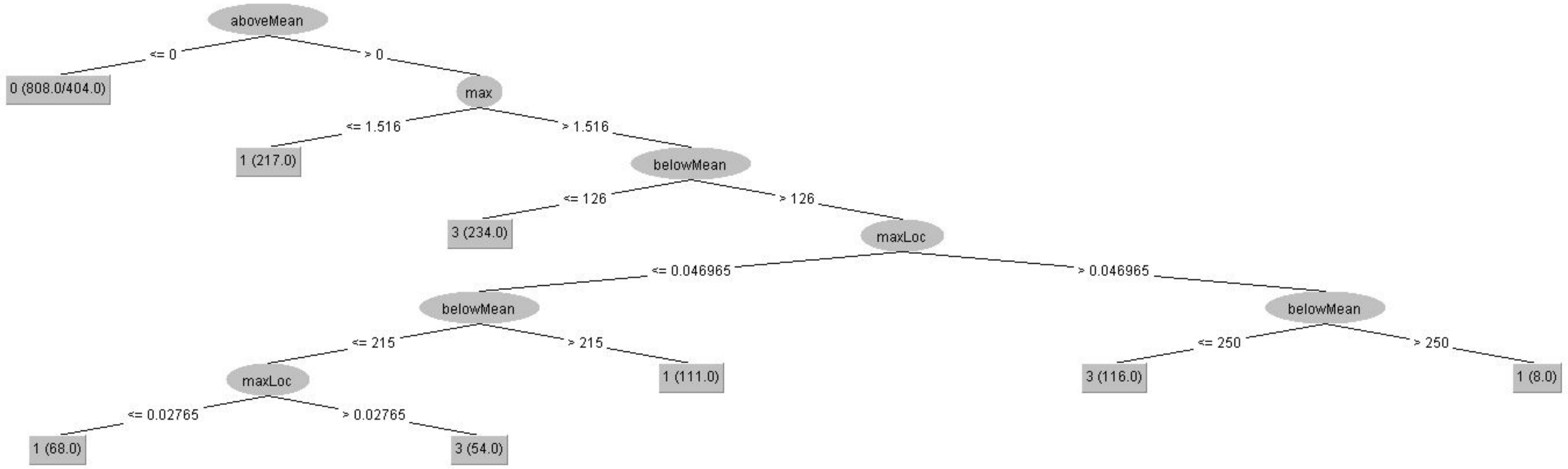
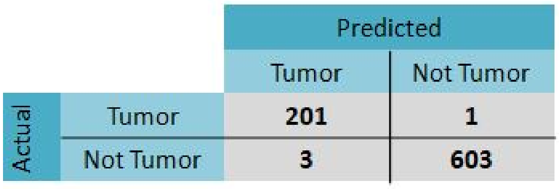


Figure 5.16: The J48 decision tree for patient 1.

Patient 2

The decision tree generated with only patient 2's data is rather short owing to the few frames in the sonograph of the patient. The tree is visualized in figure 5.18.

The classifier achieves a hight correct prediction percentage of 81. The weighted precision and recall are also high; both being 0.81. As expected, the kappa statistic is also high at 0.75. Further, the confusion matrix of the tumor class shows high true positives and tru negatives and low false positives and false negatives. The numbers are shown in figure 5.17. The weighted false positive rate is also low at 0.06.



		Predicted	
		Tumor	Not Tumor
Actual	Tumor	201	1
	Not Tumor	3	603

Figure 5.17: The confusion matrix for the tumor class with the J48 decision tree classifier on patient 2's data.

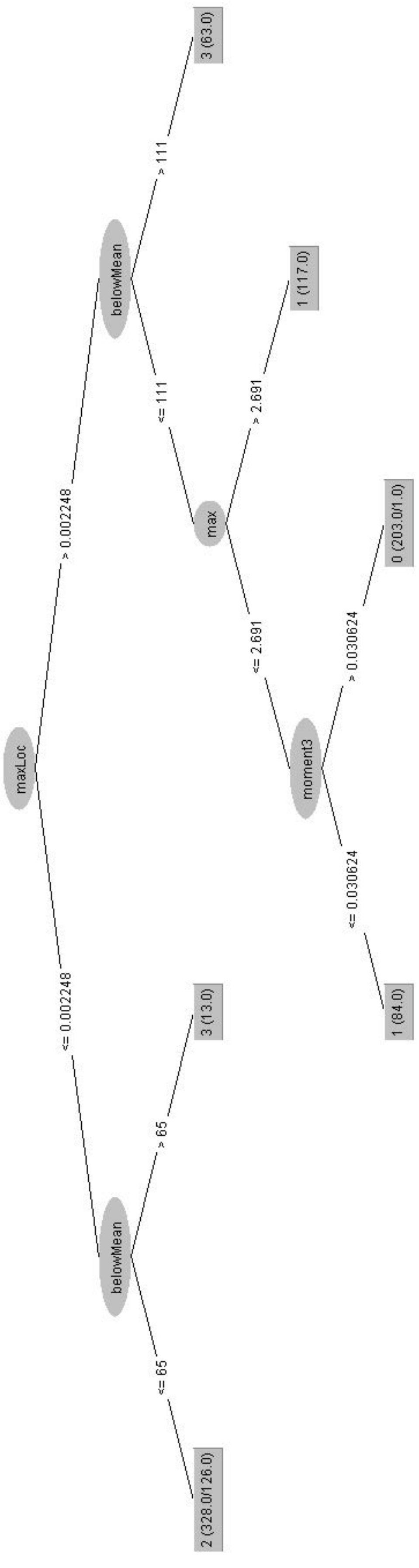


Figure 5.18: The J48 decision tree for patient 2.

Summary

For both patient 1 and patient 2, the J48 decision tree classifier does not perform very well. As shown in figure 5.19, the model trained with only patient 1 data does not identify the boundaries and wrongly classifies class 0. While on the other hand, the comboned model rightly identifies the boundary class 2 but a majority of the frame is classified as part of tumor (class1). Hence, the classification quality is poor.

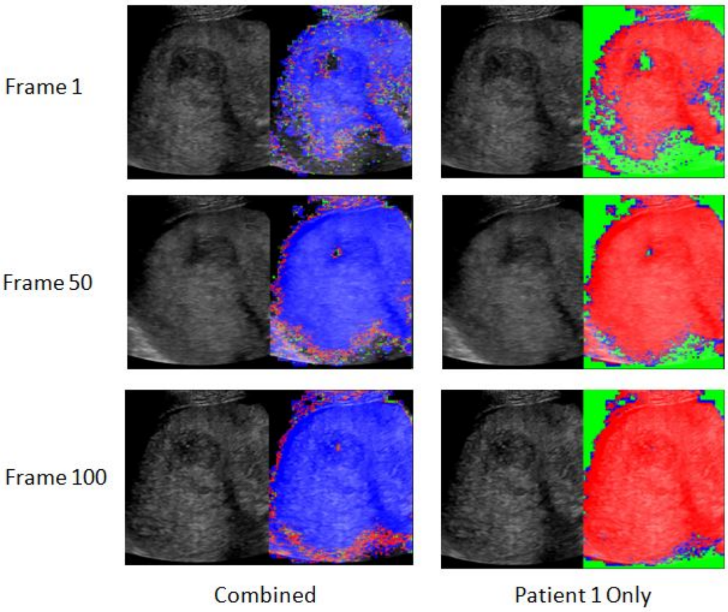


Figure 5.19: Frames of patient 1's CEUS video as segmented using the J48 Decision Tree model trained with combined data and data only from patient 1.

The model trained with the combination of data from the two patients and one trained only with patient 2's data, both identify the boundaries well but show weak classification results for the tissues. This is shown in figure 5.20 where majority of the tissue is classified as part of tumor (class 1) in blue.

The decision tree algorithm is known to overfit. Hence, it usually produces a high accuracy. This is reflected in the kappa statistic of the classifiers. Also the false positives are found to be unreasonably low.

The evaluation statistics captured during the 10FCV on the three types of training data used for the J48 decision tree classifier are summarized in table 5.4. This is an example of why we can not evaluate a solely model based on the performance statistics and that we require human evaluation in place.

Data	%Correct	%Incorrect	Kappa	Weighted AUC	WTP	WFN	WFP	WTN	WPrecision	WRecall
Combined	79.08	20.92	.72	.91	.79	.21	.07	.93	.88	.79
Patient 1	74.69	25.31	.66	.92	.75	.25	.08	.92	.75	.75
Patient 2	81.06	18.94	.75	.95	.81	.19	.06	.94	.81	.81

Table 5.4: Evaluation statistics for the J48 Decision Tree Classifier.

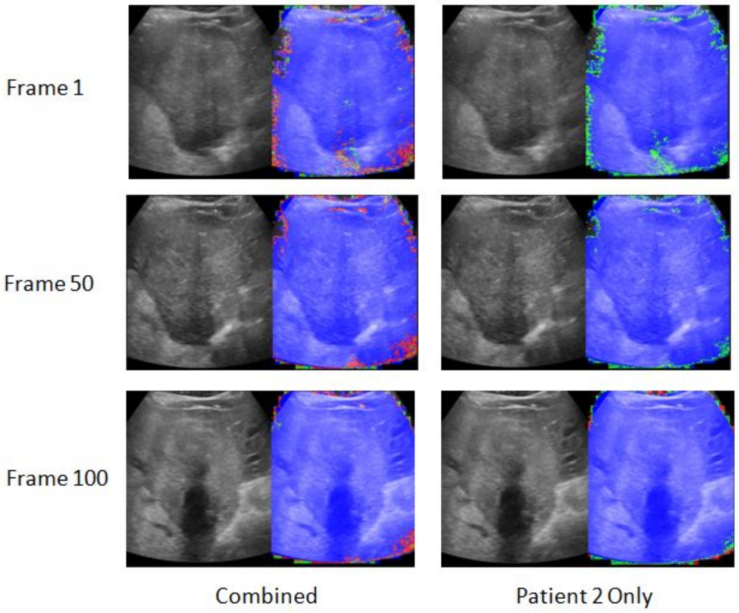


Figure 5.20: Frames of patient 2's CEUS video as segmented using the J48 Decision Tree model trained with combined data and data only from patient 2.

5.2.2.3 Random Forest

Combined

The random forest classifier trained with the combined data of two patients produces a weighted AUC of 0.84. The AUC for the tumor class is 0.75 as shown in figure 5.21. The classifier correctly classifies 68.13% instances. It has weighted precision and recall of 0.7 and 0.68 respectively. The weighted false positive rate also stays low at 0.11. The kappa statistic is found to be 0.58.

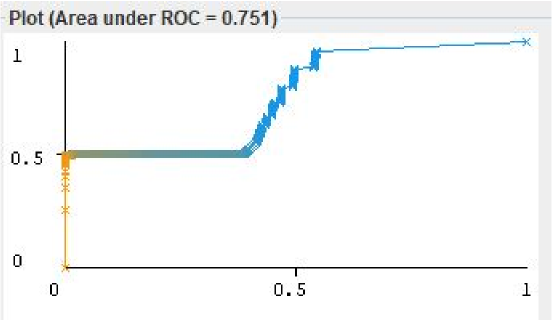
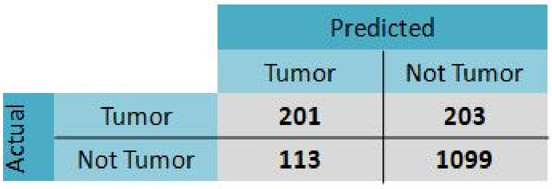


Figure 5.21: The area under ROC is 0.75 for the random forest classifier trained on data from two patients for tumor class (class 0).

The confusion matrix corresponding to the tumor class is shown in figure 5.22. We observe that the false positive rate is a low 0.09. Low false positive rates were also observed for the J48 classifier.

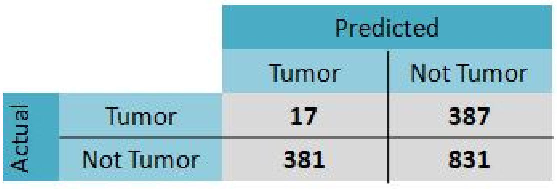


		Predicted	
		Tumor	Not Tumor
Actual	Tumor	201	203
	Not Tumor	113	1099

Figure 5.22: The confusion matrix for the tumor class with the random forest classifier on combined patient data.

Patient 1

The weighted AUC of the random forest classifier trained on patient 1's data is 0.84 while the AUC for tumor class (class 0) is 0.67. The weighted precision and recall drop to 0.52 each. As a result the kappa statistic also drops. It is found to be 0.37. The percentage of correctly classified instances also reduces to 52.5. The confusion matrix for class tumor is shown in figure 5.23. The number of false positives is seen to rise. Consequently, the false positive rate increases to 0.31.

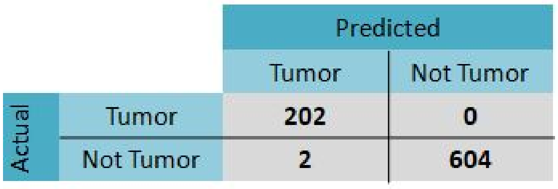


		Predicted	
		Tumor	Not Tumor
Actual	Tumor	17	387
	Not Tumor	381	831

Figure 5.23: The confusion matrix for the tumor class with the random forest classifier on patient 1's data.

Patient 2

The false positive rate is seen to be low again for the model trained with patient 2's data alone. The confusion matrix is shown in figure 5.24 and results in a false positive rate of 0.1. While the AUC for the tumor class is 1, the weighted AUC is 0.94 and the the weighted precision and recall are 0.7 each. The kappa statistic is seen to increase to 0.6. The effect of this is reflected in the percentage of correctly classified instances that raises to 70%.



		Predicted	
		Tumor	Not Tumor
Actual	Tumor	202	0
	Not Tumor	2	604

Figure 5.24: The confusion matrix for the tumor class with the random forest classifier on patient 2's data.

Summary

Figure 5.25 shows the resulting classifications from the random forest classifier trained with the combination of patient data and with only patient 1's data. The model trained on the combined data identifies the boundaries well but does not do well in identifying the tumor tissues. On the other hand, the model trained with only patient 1's data identifies few pixels of tumor class 0 correctly and again most of the frame is classified as part of tumor tissue belonging to class 1. However, the model is quite unsure about the boundary class.

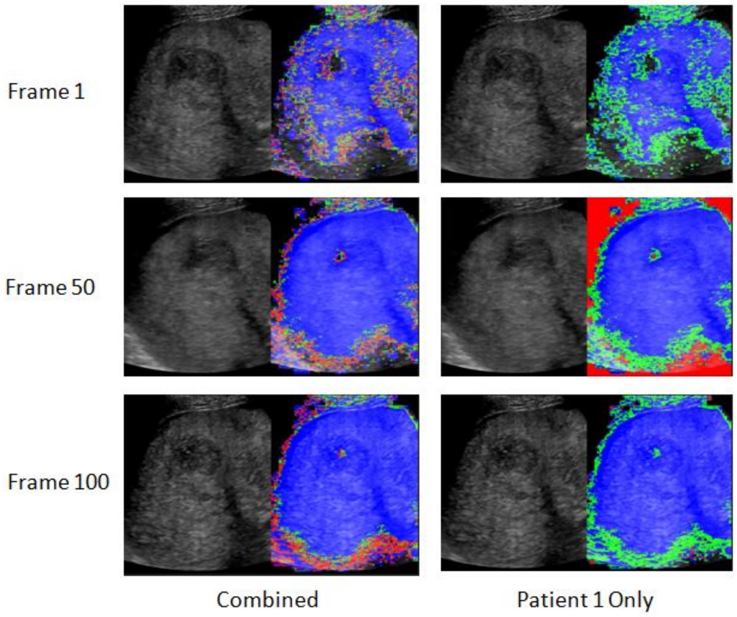
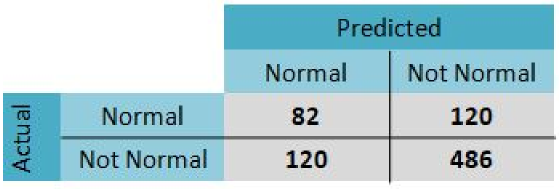


Figure 5.25: Frames of patient 1's CEUS video as segmented using the Random Forest model trained with combined data and data only from patient 1.

As for patient 2, the classification results are shown in figure 5.27. The classification is not informative as majority of the image is classified as class 1- part of the tumor tissue. The boundaries are identified much better in the combined model than in the model trained with data from patient 2 only.



		Predicted	
		Normal	Not Normal
Actual	Normal	82	120
	Not Normal	120	486

Figure 5.26: The confusion matrix for the normal tissue class - class 3, with the random forest classifier on patient 2's data.

The confusion matrix of the model trained only on patient 2's data has a high number of true positives and true negatives. The number of false positives and false negatives are low. This directs us to assuming that the predictions should be good but as seen, this is not the case. This can be understood with the confusion matrix of class 3 shown in figure 5.26. The false negative rate is high at 0.59. This means that majority of pixels not belonging to class 3 are wrongly classified as being class 3.

The 10FCV evaluation statistics are summarized in table 5.5.

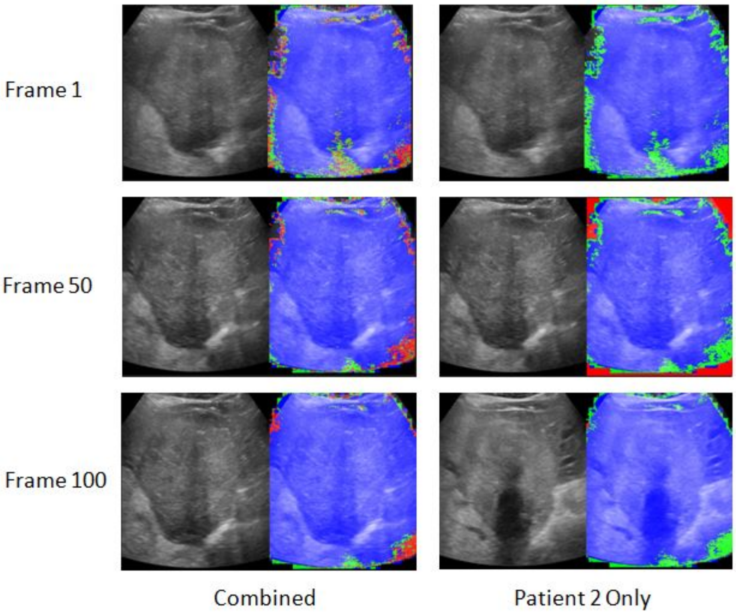


Figure 5.27: Frames of patient 2's CEUS video as segmented using the Random Forest model trained with combined data and data only from patient 2.

Data	%Correct	%Incorrect	Kappa	Weighted AUC	WTP	WFN	WFP	WTN	WPrecision	WRecall
Combined	68.13	31.87	.58	.84	.68	.32	.11	.89	.7	.68
Patient 1	52.48	47.52	.37	.84	.52	.48	.16	.84	.52	.52
Patient 2	70.05	29.95	.6	.94	.7	.3	.1	.9	.7	.7

Table 5.5: Evaluation statistics for the random forest Classifier.

5.2.2.4 Sequential Minimal Optimization

Combined

The classifier built with the combined data of two patients has a AUC of 0.69 for tumor class 0 as shown in figure 5.28 while the weighted AUC is 0.8. The weighted false positive rate is low at 0.13. The model classifies instances correctly in 62.4% cases. The kappa statistic is quite average at 0.5. The weighted precision and recall are 0.65 and 0.62 respectively.

The confusion matrix corresponding to the tumor class 0 is shown in figure 5.29. The false

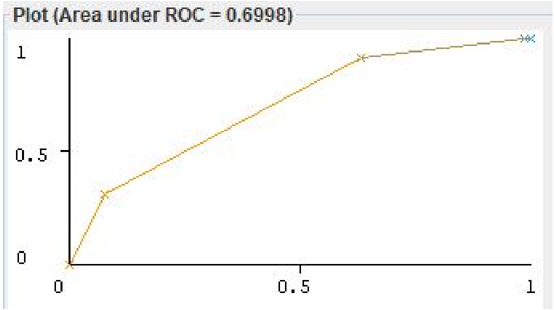
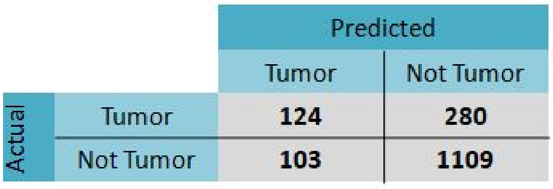


Figure 5.28: The area under ROC is 0.69 for the tumor class 0 using the SMO classifier trained on data from two patients.

positive rate looks good at 0.08. However, the false negative rate is quite high a 0.69 indicating that the model is not sure about the actual tumor class and predicts it as non-tumor many times.

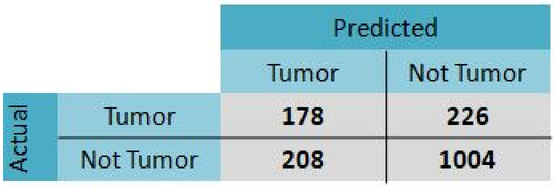


		Predicted	
		Tumor	Not Tumor
Actual	Tumor	124	280
	Not Tumor	103	1109

Figure 5.29: The confusion matrix for the tumor class with the SMO classifier on combined patient data.

Patient 1

As for the model trained with data only from patient 1, a weighted AUC of 0.89 is observed. The AUC for the tumor class 0 is 0.82. The kappa static increase marginally to 0.57 and the model classifies 67.88% instances correctly. Both weighted false positive and false negative rates drop. The weighted precision and recall increase to 0.68 each. The confusion matrix of the tumor class 0 shown in figure 5.30 indicates that the false positive rate has increased to 0.17 and the false negative rate has dropped to 0.56.

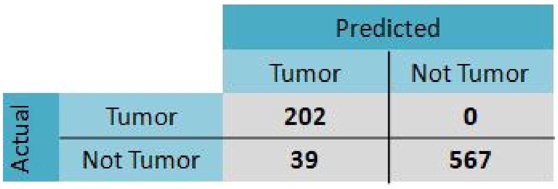


		Predicted	
		Tumor	Not Tumor
Actual	Tumor	178	226
	Not Tumor	208	1004

Figure 5.30: The confusion matrix for the tumor class with the SMO classifier data from patient 1.

Patient 2

The classifier trained with patient 2's data is found to correctly classify 77.35% instances. This is reflected in the kappa statistic which is 0.7.



		Predicted	
		Tumor	Not Tumor
Actual	Tumor	202	0
	Not Tumor	39	567

Figure 5.31: The confusion matrix for the tumor class with the SMO classifier data from patient 2.

The weighted precision and recall also increase to 0.82 and 0.77 respectively. The weighted AUC is 0.89 while the AUC for tumor class 0 is a whooping 0.97. The weighted false negative and false positive rates are also low at 0.23 and 0.08 respectively. The confusion matrix of tumor class 0 shown in figure 5.31 shows that actual tumor class was always predicted correctly. Hence, the false negative rate is 0. However there were a few non-tumor instances predicted as tumors. This false positive rate was low at 0.06.

Summary

The overall 10FCV evaluation metrics for the three training cycles are shown in table 5.6. The AUC for the tumor marked as class 0 is seen to be higher for models trained on individual patients than the combined model. This is also reflected in the segmented images that are generated in figures 5.32 and 5.33. The boundaries are identified better in the model with combined patient data.

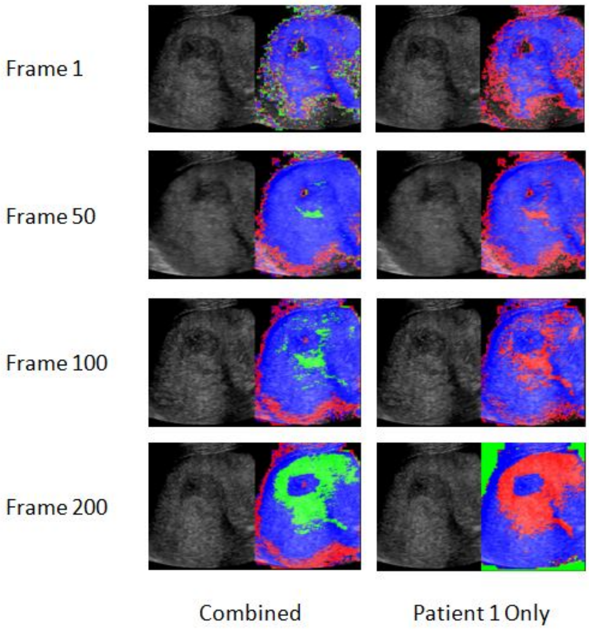


Figure 5.32: Frames of patient 1's CEUS video as segmented using the SMO model trained with combined data and data only from patient 1.

Data	%Correct	%Incorrect	Kappa	Weighted AUC	WTP	WFN	WFP	WTN	WPrecision	WRecall
Combined	62.44	37.56	.5	.8	.62	.38	.13	.87	.65	.62
Patient 1	67.88	32.12	.57	.89	.68	.32	.11	.89	.68	.68
Patient 2	77.35	22.65	.7	.89	.77	.23	.08	.92	.82	.77

Table 5.6: Evaluation statistics for the SMO Classifier.

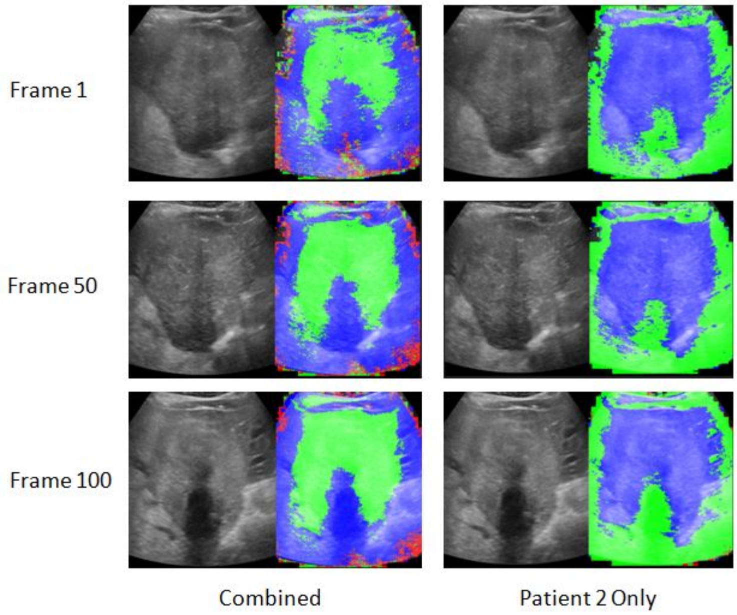


Figure 5.33: Frames of patient 2's CEUS video as segmented using the SMO model trained with combined data and data only from patient 2.

5.2.2.5 Similarity based - k Nearest Neighbours

Combined

The tumor class (class 0) shows an AUC of 0.72 as shown in figure 5.34. The weighted AUC of the classifier on the combined data of the two patients is 0.77.

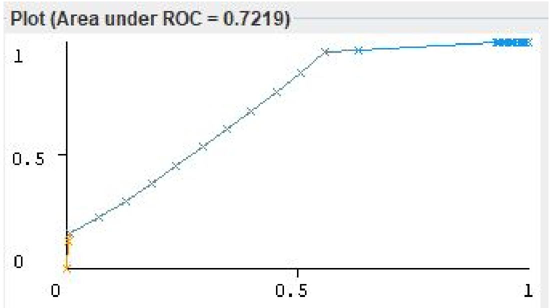
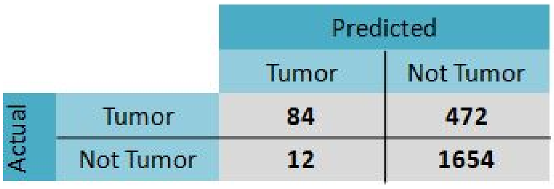


Figure 5.34: The area under ROC is 0.72 for the tumor class 0 using the kNN classifier trained on data from two patients.



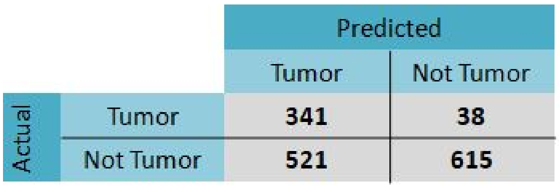
		Predicted	
		Tumor	Not Tumor
Actual	Tumor	**84**	**472**
	Not Tumor	**12**	**1654**

Figure 5.35: The confusion matrix for the tumor class with the kNN classifier on combined patient data.

The weighted false positive rate is low at 0.14. However, the weighted false negative rate is is relatively high at 0.41 implying that the classifier is poor in identifying tumor tissues. The weighted precision and recall ate 0.76 and 0.59 respectively. The confusion matrix for the tumor class 0 is shown in figure 5.35.

Patient 1

In terms of the AUC of the tumor class, the model trained on patient 2 does not show major difference. The AUC of class 0 is 0.73. The weighted AUC also increases to 0.79. The weighted false positive and weighted false negative rates also do not change a lot and are 0.17 and 0.5 respectively. The confidence matrix of tumor class 0 observed is shown in figure 5.36.

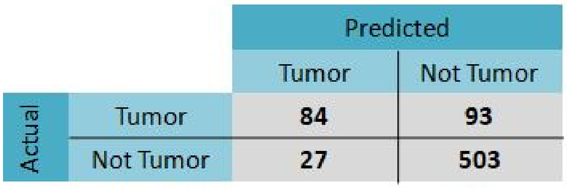


		Predicted	
		Tumor	Not Tumor
Actual	Tumor	341	38
	Not Tumor	521	615

Figure 5.36: The confusion matrix for the tumor class with the kNN classifier data from patient 1.

Patient 2

As for patient 2, the confusion matrix for the tumor class (class 0) is shown in figure 5.37.



		Predicted	
		Tumor	Not Tumor
Actual	Tumor	84	93
	Not Tumor	27	503

Figure 5.37: The confusion matrix for the tumor class with the k-NN classifier data from patient 2.

Summary

Figure 5.38 shows the results of classification by the k-NN model trained with combination of patient data and only patient 1's data. The combined model performs better at identifying the boundaries but does poorly at classifying tumors.

As for patient 2, the combined model again identifies boundaries well and also does well on the tumor tissues. The individual classes are very well classified by the k-NN classifier on patient 1. Notably, the model trained on patient 1's data only, produces better results for the test frames in this case. The classification outputs are shown in figure 5.39.

The summary statistics from the 10FCV of the different training data are shown in table 5.7. The correctly classified instances in the first two cases is 50%. This indicates that the classifier simply guesses and is rather poor. This is seen to improve when trained only with patient 1's data but is still low in comparison to other classifiers.

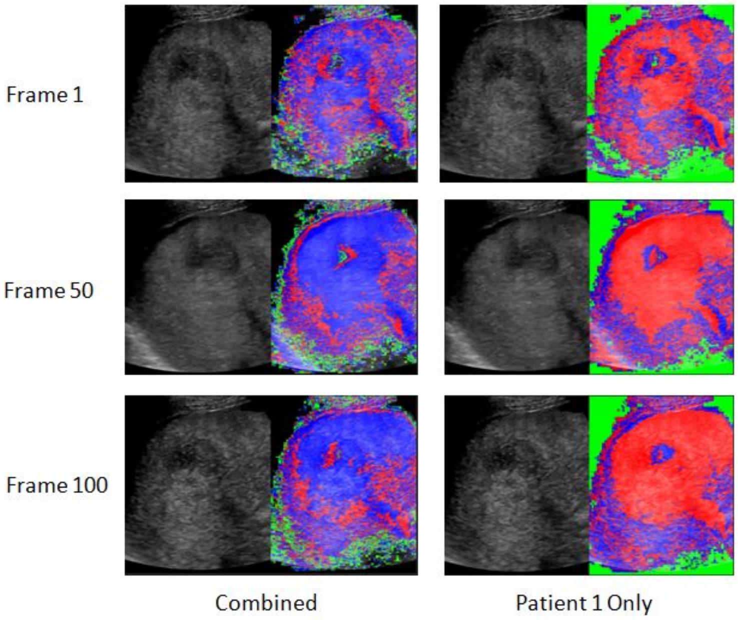


Figure 5.38: Frames of patient 1's CEUS video as segmented using the k-NN model trained with combined data and data only from patient 1.

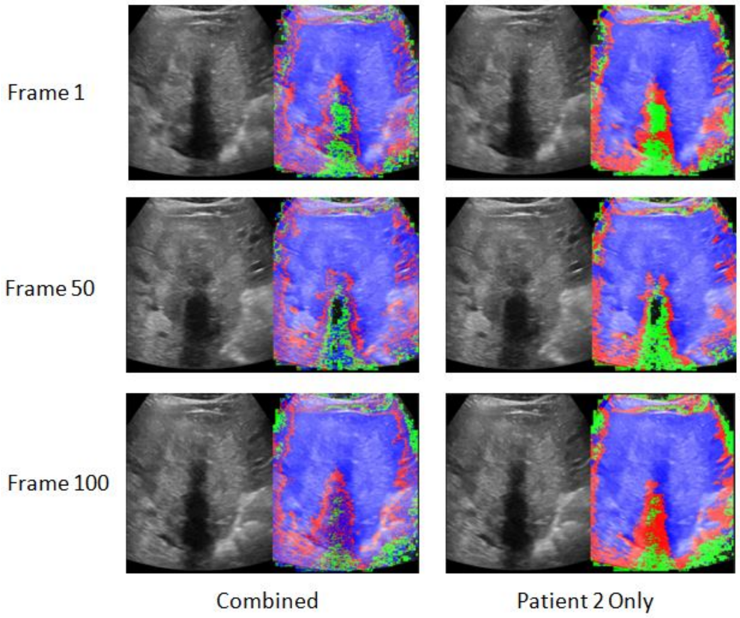


Figure 5.39: Frames of patient 2's CEUS video as segmented using the k-NN model trained with combined data and data only from patient 2.

Data	%Correct	%Incorrect	Kappa	Weighted AUC	WTP	WFN	WFP	WTN	WPrecision	WRecall
Combined	58.82	41.18	.45	.77	.59	.41	.14	.86	.76	.59
Patient 1	59.8	40.2	.46	.79	.6	.4	.13	.87	.67	.6
Patient 2	59.69	40.31	.46	.75	.6	.4	.13	.87	.71	.6

Table 5.7: Evaluation statistics for the k-NN Classifier.

5.2.2.6 Discussion

We trained classifiers on the CEUS data of two patients both individually and using a combination of the two. Various evaluation metrics were discussed in the previous sections. We looked at the weighted metrics and also the metrics pertaining specifically to the tumor tissue (class 0). We now rank the classifiers based on the weighted AUC, tumor class AUC and percentage of correctly classified instances. Table 5.8 ranks the classifiers trained on the combination of data from the two patients.

Algorithm	AUC - Tumor Class	Weighted AUC	% Correct
j48	0.87	0.91	79.08
Random Forest	0.75	0.84	68.13
Sequential Minimum Optimization	0.69	0.8	62.44
k-NN	0.72	0.77	58.82
Naive Bayes	0.62	0.75	49.69

Table 5.8: Classifiers trained on combined patient data ranked by performance after 10FCV.

As for the models trained on individual patients, J48 decision tree and SMO are seen to have the best performance in terms of the weighted AUC. However, the classification results produced look much better for SMO classifier for both patients. The k-NN classifier performs specially well for patient 1.

Algorithm	AUC- Tumor Class	Weighted AUC	% Correct
Sequential Minimum Optimization	0.82	0.89	67.88
j48	0.83	0.92	74.69
k-NN	0.72	0.79	59.8
Random Forest	0.67	0.84	52.48
Naive Bayes	0.81	0.85	50.68

Table 5.9: Models trained on patient 1 ranked by performance.

Algorithm	AUC- Tumor Class	Weighted AUC	% Correct
j48	0.99	0.95	81.06
Sequential Minimum Optimization	0.97	0.89	77.35
Random Forest	1	0.94	70.05
Naive Bayes	0.98	0.83	68.19
k-NN	0.71	0.75	59.69

Table 5.10: Models trained on patient 2 ranked by performance.

When training, it is observed that models with the combined data do not perform much worse than those trained on individual patients. Given the case, it is better to train models on combined patient data as it is more generalized. However, the data preparation step would then need to include and additional step for the combination. Figures 5.40 and 5.41 shows the classification results on selected frames for patient 1 and patient 2, respectively, produced from the models trained with combination of data from the two patients.

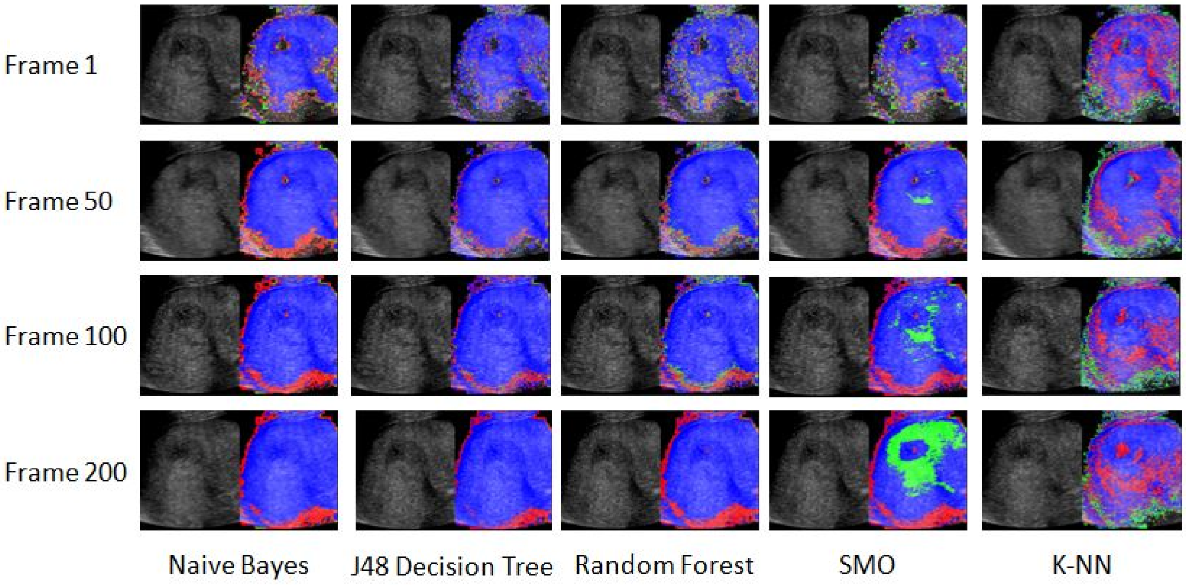


Figure 5.40: Frames of patient 1's CEUS video as classified by various classifiers rained on combined data from two patients.

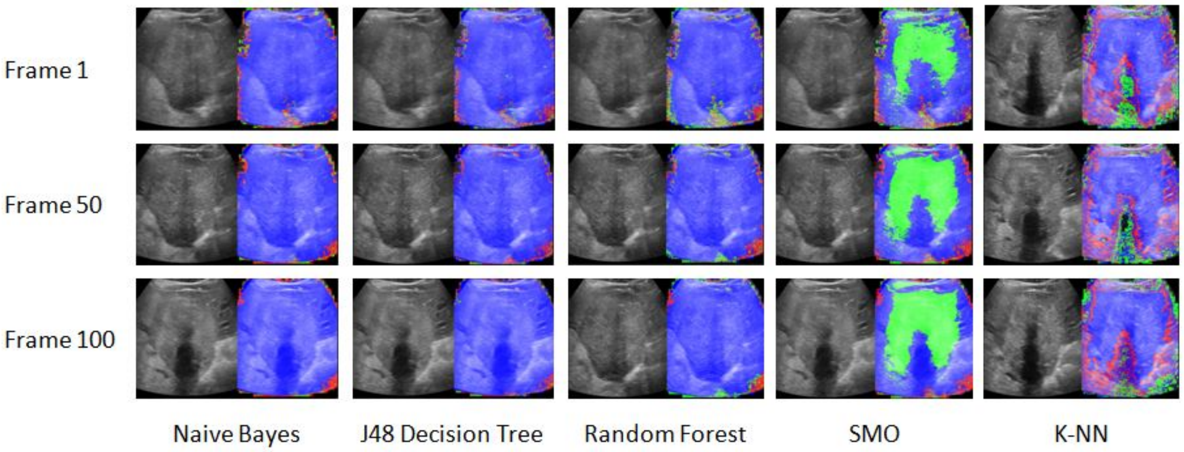


Figure 5.41: Frames of patient 2's CEUS video as classified by various classifiers rained on combined data from two patients.

5.3 Deep Learning Approach

5.3.1 Evaluation Setup

As in the time-series approach, we also combined videos of patients for the convolution neural network based learning. In this experiment, the data from 4 patients was used. The shortest among the 4 sonograph videos had 163 frames. We truncate the other videos to select the first 163 frames from them. Hence, the total data size is 652 frames. However, since hand-labelling 652 images is tedious we chose a total of 193 images from the first 163 frames of the 4 videos. These were labelled and constituted the training set along with their augmentations as described in section 4.3.2. To train the U-net, masks are created based on the information received from the physicians. The masks vary per frame for each patient. However, they are exactly the same for the US and CEUS components in a single frame. Figure 5.42 shows one frame from each patient, the corresponding annotated image and the mask generated with that information.

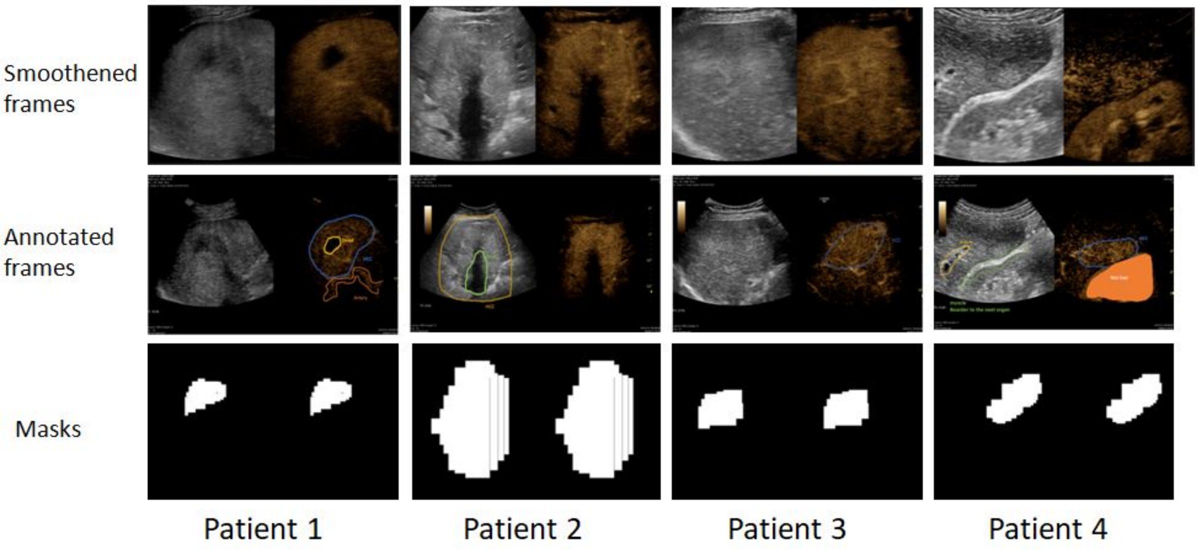


Figure 5.42: Examples of masks generated for the four patients on selected frames.

Further, the pixel values were normalized to lie in the range $[0,1]$ and the data was divided into training and test set. The training images were augmented as explained in section 4.3.2.

5.3.2 Evaluation Results

The U-net was trained for 5 epochs of 2000 steps per epoch. We ran the network on CPU with 32 GB RAM. The training lasted 5 days. The progress of the accuracy over the epochs trained is plotted in figure 5.43. The trained U-net displayed and accuracy of 97.13

Also, the binary cross entropy loss used was monitored throughput the training phase and has a final value of 0.09.

Accuracy is an evaluation metric with high variance. Hence, we can not base the goodness of the network solely on accuracy. Therefore, the intersection over union (IOU) score is also calculated for the U-net trained trained. The IOU score achieved is **95.89%**.

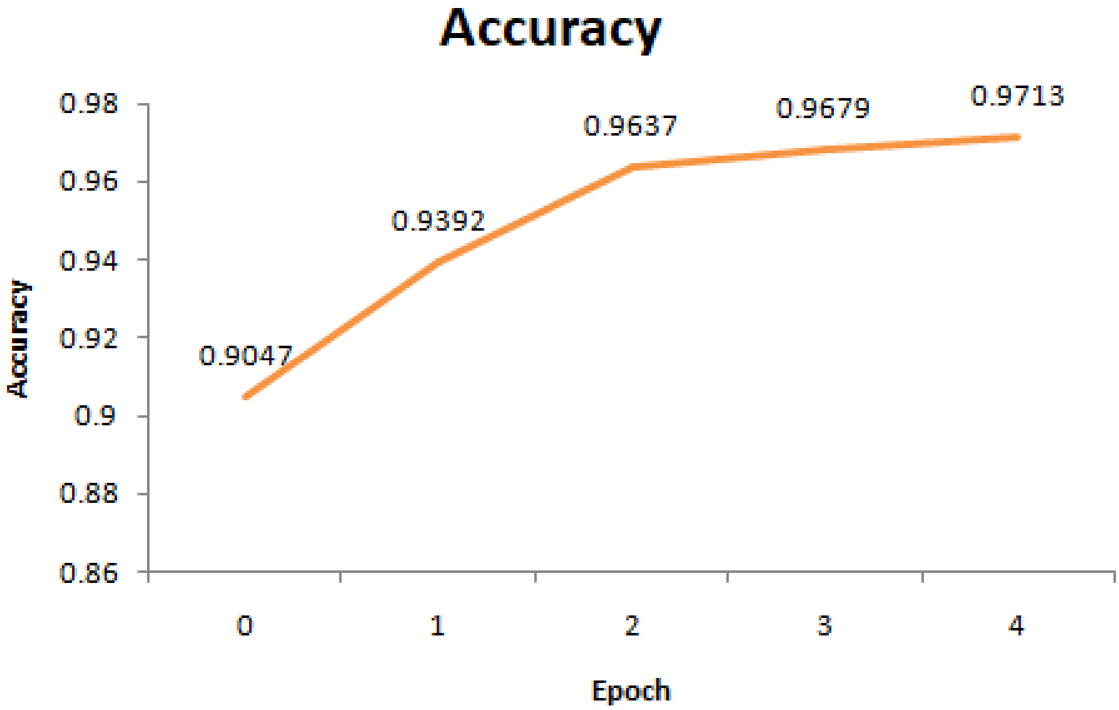


Figure 5.43: Accuracy of the network during training. The final accuracy is

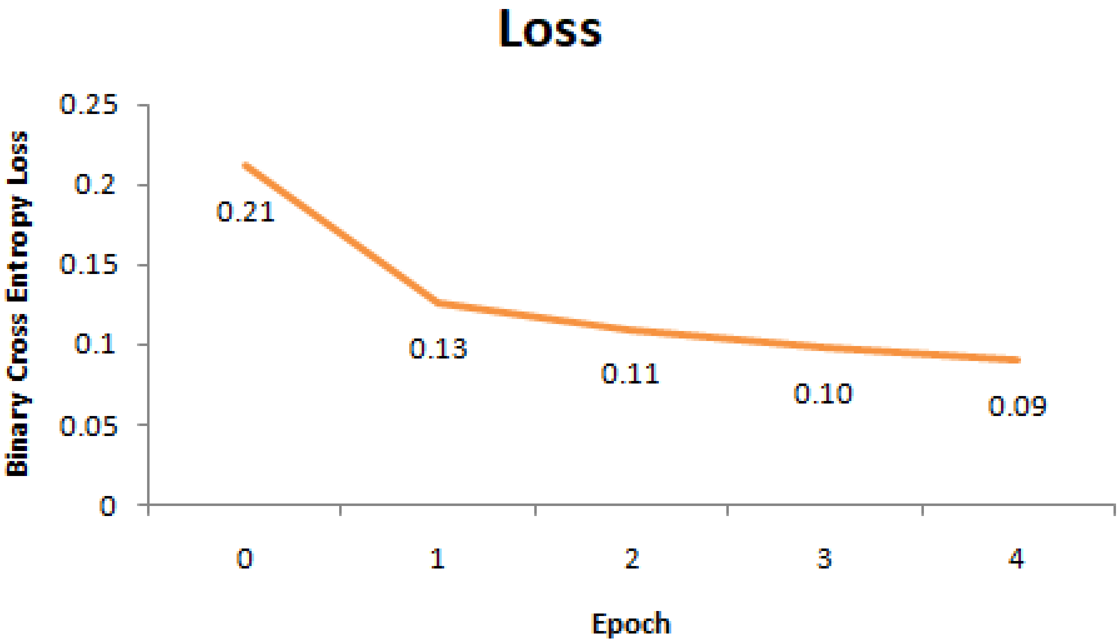


Figure 5.44: Binary crossentropy loss of the network during training. The final loss is

5.3.2.1 Visualizing results

The U-net was trained with masks corresponding to each frame. Hence, when the test frames were run, the output of the U-net was masks. These masks were then overlaid on the test frames to generate the final output. Some examples of final results are shown in figure 5.45.

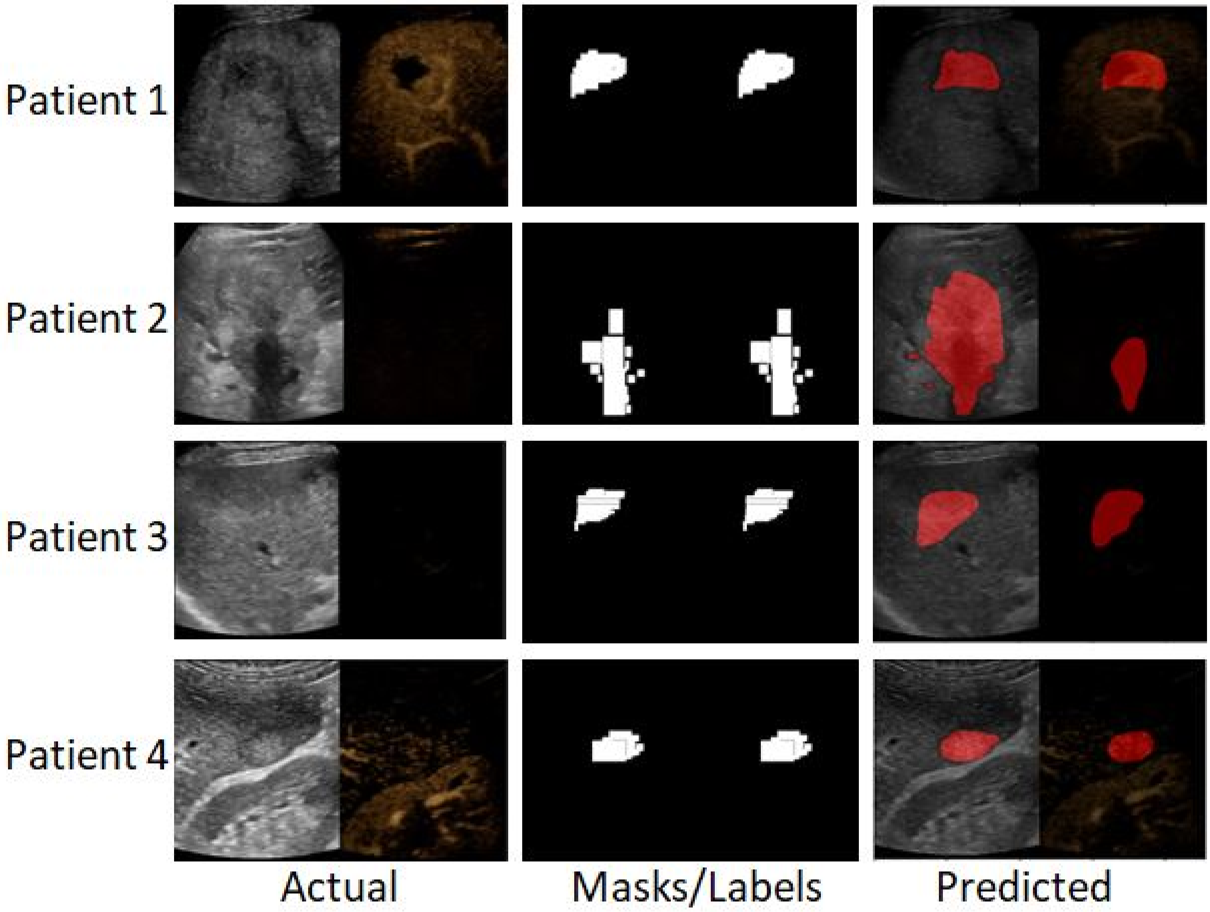


Figure 5.45: Four actual test frames, marked labels and final results from the U-net for the four patients.

Since the data was labelled such that the masks on the CEUS and US side of a frame were always the same, in some cases when there is no CEUS image, masks are still marked. This is beneficial as the network inherently learns features even when the CEUS parts are absent. The result of this is seen in patient 2's predictions in figure 5.45. The CEUS side is only beginning to appear but the network is already able to identify major dead tissue cells.

5.4 Summary

The various machine learning techniques applied have been extensively compared and discussed in this chapter. We looked at various evalion metrics to compare the performance of the models trained. In terms of observed predictions, SMO and k-NN are found to be the top performers. They also show an high AUC close to 1. However, in terms of traditional metrics like accuracy and percentage of correctly predicted instances, tree based algorithms- J48 and random forest show better numbers. This is explained by their nature to overfit training data.

The U-net trained is seen to produce good results and it identifies tissues not readily visible to the human eye with its extensive feature generation. It is observed that the deep learning method produces excellent results with fewer data having more variability among patients. We achieved an IOU of **95.89%** on the CEUS videos of 4 patients combined.

In comparison, the results of the deep learning model are found to be much better than the machine learning approach based on time-series. However, the good results are obtained at the cost of long training time and availability of decent hardware to run the network. The training time of the machine learning algorithms is shorter than the U-net.

5.5 Future Work

Although the analysis and algorithms used have been discussed in depth, we believe there is a lot of scope for further work on this topic.

In terms of the data, in the current experiment, we worked with the CEUS scans of 5 patients. However, we could broaden the scope and include scans from more patients to increase the range of training domain even more. This would require labelling a larger set of frames; translating directly to more man hours. Further, at present we only collect data from a single ultrasound machine - GE Healthcare Logiq E9, using one particular contrast agent - SonoVue. One might think using data from multiple machines would also promote the robustness of our scheme. However, we would not claim so because

1. each machine uses ultrasound waves of different frequencies and also the inbuilt logic and hardware affect the final scan, for instance, in terms of output color scheme.

2. every contrast agent enhances the tumors and normal tissues separately but the duration and intensity of enhancement vary within contrast agents.

Hence, we would want to increase the number of patients being considered while training but restrict to one machine and one contrast agent. However, the process of data collection and labelling can be made decentralized across hospitals to further promote research on the topic. The positive effects of creation of a good data source, especially for images has been well established with the creation and popularization of the ImageNet data source by Jia Deng et al. [44] in 2009.

With respect to the algorithms experimented with, in this thesis we only study one deep neural network architecture - U-Net's performance on the dataset. There are other architectures like recurrent neural networks (RNN) a combination of RNN and CNN [89] that can be tested on the dataset. Furthermore, the bottleneck of labelling data could be reduced by evaluating a unsupervised learning scheme using long short term memory (LSTM) networks [75].

6
State of the Art

6.1 Computer Aided Diagnostics

Shamir et al. [69] analyze available pattern recognition (PR) approaches for microscopy image analysis. Instead of concentrating on a particular organ or imaging technique used, they identified 4 general steps that most PR approaches have in common. As a first step, these algorithms select a region of interest (RoI). This can be done manually but this often introduces bias and leads to inconsistency. Hence, automatic RoI selection is preferred in most cases. Other widely used methods for ROI detection include global thresholding [93], watershed algorithms [87], [52], [64], model-based segmentation [30], and contour methods [88]. Li et al. [50] use automatic edge detection to segment regions of interest.

As a second step, the algorithms summarize pixel data to extract more information on image content using feature extraction. These algorithms are usually very general in that they can operate on any set of pixels without specifying any parameters. However, since we are focused on CEUS scans of the liver in this work, we add features specific to this imaging technique too.

The next step analyzed by Shamir et al. [69] is Feature Selection and Classification. As for Feature Selection, it can be done using *filters*, *wrappers* or *embedded techniques*. In filtering, statistical methods are used to compare features and select the most discriminating ones. The feature selection is blind to the classification algorithm to be applied. On the other hand, in wrapping features are selected based on their performance in the classifier being applied. Hence, the feature selection is aware and dependent on the classifier. In the embedded feature selection technique, feature selection is done as part of the classification process by the classifier. This is done in decision tree classifiers and is less computationally intensive than wrappers. The paper however, does not present an elaborate analysis on the landscape of classifiers applicable but instead concentrates on the Support Vector Machine (SVM) classifier. The authors found that for the microscopic image analysis, the number of classes are usually in the order of a few thousands and SVM is the only classifier that is moderately tested in literature. However, at the macroscopic analysis of biological images, classifiers including Random Forests [38], Bayes Classifier [35] and Artificial Neural Networks [72] have been used.

The final step of the PR approach to image analysis is interpreting the classification output and evaluating the goodness of the applied classification algorithm. Accuracy of a classifier, defined as the ratio of correctly classified test images by the total number of images tested, is one metric to measure its goodness. However, depending on the set of images used for training and testing, accuracy can be biased. To avoid this, cross-validation is used and the training and testing sets are randomly shuffled and the classifier accuracy is measured in multiple runs. Still, Shamir et al. [69] rightly argue it might be possible that the accuracy does not have any biological meaning. More often than not, labeling of classes is manual. This is a major cause of bias. Further, the equipment used for imaging also adds some bias. In most cases, labeling is aided by the features generated or by looking at a *processed* image rather than raw. Therefore, the artifacts influence class assignment while labeling. Hence, it is possible that the classifier has very high accuracy but this is only due to the way the training data is labeled. It might be far away from the biological truth. It is therefore important to ensure that the data collection and labeling is unbiased. Another measure of classifier goodness is the *confusion matrix* or the *error matrix*. The matrix summarizes the True Positives and True Negatives, that is the number of times a test image's class was correctly predicted. Additionally, it provides information about the False Positives and False Negatives. Each of which happens when the test image belongs to one class and the classifier predicts it to belong to another class. The dimension of the confusion matrix reflects the number of classes in the classification problem. A binary classification problem has a 2 x 2 confusion matrix. A *n*-class classification problem has a *n x n* confusion matrix. The confusion between a pair of classes also gives an estimate about their similarity. This is also used for biological analysis [69].

Partly, we also follow this 4 step approach to classify normal tissues and lesions within CEUS data collected. RoIs are identified by the radiologists and tissues are labeled by them. Based on a considerably sized training set, we generate feature vectors per pixel per frame. We test tree based classifiers - J48 Decision Tree, Hoeffding Tree and Random Forest, with embedded feature selection among other classifiers on the training sets. The other classifiers we test include Naive Bayes, Sequential Minimal Optimization (SMO) and k- Nearest Neighbours (kNN). We perform 10 fold cross validation on all models trained to evaluate their performance.

Ghose et al. [38] propose an approach to prostate gland segmentation based on building multiple mean parametric models derived from principal component analysis of shape and posterior probabilities in a multi-resolution framework. They work with 126 ultrasound images of the prostate gland taken from different positions of the transducer. They use supervised learning to build a Random Forest classifier to determine the posterior probability of a pixel being prostate. As a second step, they build a statistical shape and appearance model. On this front, the authors build a point distribution model (PDM) by equal-angle sampling of the prostate contours to model shape and use the active appearance model (AAM) from [3] for appearance modeling. They apply principal component analysis of the PDM and AAM to identify principal modes of variation in shape and appearance. Motivated by central limit theorem which states that a non-Gaussian distribution can be better approximated by multiple Gaussians, the authors use multiple mean Gaussian models of shape and appearance. The authors use prostate segmentation evaluation metrics - Dice Similarity Coefficient (DSC) , 95% Hausdorff Distance (HD), Mean Absolute Distance (MAD), specificity and sensitivity to evaluate their method. In this work, for the machine learning experiments, instead of an explicit shape detector, edge detection is leveraged. Canny edge detection algorithm is chosen because it also suppresses noise. This algorithm is applied at the training step and the edge information is then a part of the feature vector used for classification. In the deep learning part of our experiment, all feature generation and selection is embedded in the convolution network architecture used. Like Ghose et al. [38] we also use specificity and sensitivity to evaluate our models. In addition we use the area under ROC and intersection over union (IOU) metrics too.

Shiraishi et al. [72] explain a non-grid image warping technique for temporal subtraction of consecutive scan images. Temporal subtraction leads to improved identification of anomalies. An example is illustrated in figure 6.1. The authors built a University of Chicago approved CAD system allowing radiologists to generate temporal subtraction of two chosen images. This was integrated with the picture archiving and communications system (PACS) being used. The overall scheme of the implemented system is described in figure 6.2. The authors worked with the CT scans of patients repeated successively over periods of time. They developed non-linear image warping for the non-rigid image matching needed. This was done to accommodate patient movement or equipment misplacement resulting in changed angles or positions of subject in the resulting images. They marked regions of interest (RoI) based on tissues being observed and shifted the x,y coordinates of the image based on the cross-correlation values of the RoI in the comparison source image. However, in this thesis, we work with CEUS videos of patients. A patient video consists of upto 1147 frames. Hence generating this non-linear warping for every frame in two consecutive videos given the RoI shift is not feasible. Moreover, we currently do not have the consecutive scan data for patients available at hand.

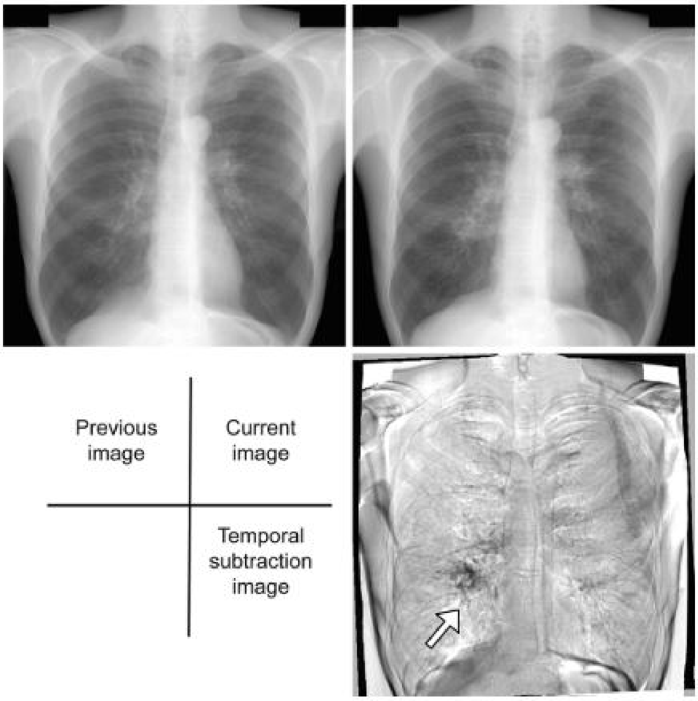


Figure 6.1: Temporal subtraction applied to the chest radiographs aids identification of a lung mass progression [72]

It is found that Artificial Intelligence through CAD is rapidly disrupting the field of medical analysis [33]. It has been studied with multiple imaging techniques like Ultrasound, CT, MRI, CEUS etc,. and also for various organs including lungs by Shiraishi et al. [72], prostate by Gelet et al. [37], breast by Finette et al. [35], liver by Lerski et al. [48] and Lerski et al. [49]. However, no record of a large scale successful implementation have been found. Additionally, no resources could be found on the video analysis of ultrasound or CEUS.

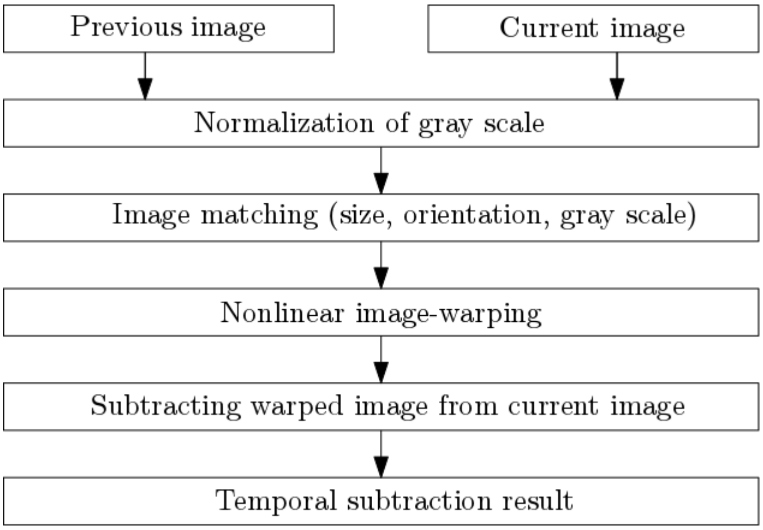

Previous image
Current image
Normalization of gray scale
Image matching (size, orientation, gray scale)
Nonlinear image-warping
Subtracting warped image from current image
Temporal subtraction result

Figure 6.2: Temporal subtraction scheme used in [72]

Although there is quite some literature available on image classification, the evaluation is usually done on in-house data sets [31], [74] . Lin et al. [51] propose a large scale image classification technique built on Hadoop using the SVM classifier on the largest available image dataset- Imagenet. They are able to achieve state-of-the-art performance on the ImageNet 1000-class classification, with 52.9% in classification accuracy and 71.8% in top 5 hit rate. Shiraishi et al. [72] use a data set of 58 consecutive bone scans to evaluate their temporal subtraction approach. Ghose et al. [38] use a total of 125 images comprising of base, central and apex zone images of the prostate gland to evaluate their supervised learning framework for automatic prostate segmentation in utrasound images. On the Microscopy analysis front, Guerra et al. [40] uses a database of 128 pyramidal cells and 199 interneurons from mouse brains to perform a comparison between supervised and unsupervised classification of neuron cell types. 380+ thermal images were used to compare different compression methods by Schaefer et al. [67]. To provide a scale for comparison, Imagenet consists of more than 14 million images. Hence, the availability of a centralized large scale medical image dataset in general and for specific organs of interest is scarce and limits application and validation of various image classification and segmentation techniques for systems like CAD.

6.2 Image Segmentation and Classification

Image segmentation plays a fundamental role in understanding image content for searching and mining in medical image archives and automated segmentation also aids diagnosis. Image segmentation is the task of grouping similar *regions* of an image together. Usually these regions are defined by the *pixels* constituting the image. Hence, pixels that are *similar* are grouped in one segment. Image Classification additionally assigns labels to identified *similar* pixels.

Sharma and Aggarwal [70] work with MR and CT images with the aim of segmentation and classification for aiding study of anatomical structure, identifying region of interests (eg. lesions) and measuring tissue volume to estimate tumor growth in order to help in treatment planning. The authors use different methods for segmentation and also list out the various possible methods for classification.

For image segmentation, the authors work with gray scale features only. They use features generated *, per image,* using each of the below listed techniques *individually*. Figure 6.3 shows the results of the segmentation techniques listed.

◇ Pixel Amplitude Histogram: A histogram of all pixels in the image with their corresponding values is built. *Similar* pixels are placed together in bins and a threshold is applied to detect a region. It is particularly suitable for an image with region or object of uniform brightness placed against a back ground of different color. The results of this technique are highly dependent on set threshold.

◇ Edge Detection: The image is segmented by the detection of edges or borders present. Again thresholds are set to define *edge-strength* based on number of segments expected. The limitations of this technique include its strong fluctuations with noise and misleading segmentation caused by weak or fake edges detected.

◇ Region Based: In this method seeds are placed and *regions* consisting of *"similar"* pixels are built around the seeds. The definition of *"similarity"* can be implementation specific. The drawback of this technique is that the image maybe over or under segmented. The authors find that one way this challenge can be rectified is by combining region based and edge detection based techniques.

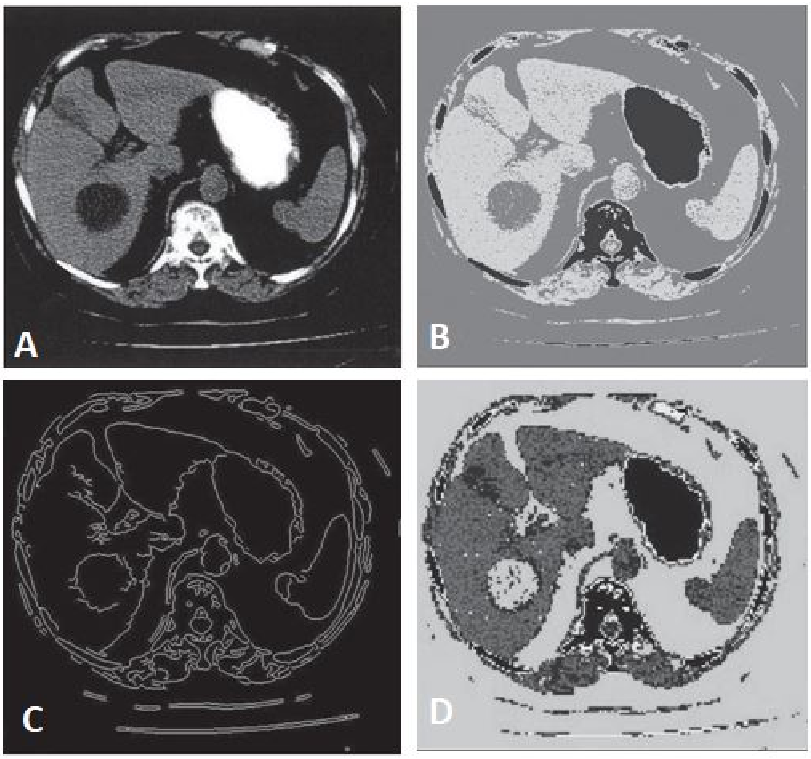


Figure 6.3: The A) Original Image of an Abdomen CT and its segmentations based on B)Pixel thresholding C)Edge Detection and D)Region seeds. [70]

Although the paper focuses on segmentation, Sharma and Aggarwal [70] identify what they call as *textural features* which also aid in classification. Texture features include not only pixel based values such as intensity, brightness etc,. but also its spatial arrangement in the image. That is, it combines the tone and structure information into one called *textone*. Two approaches the authors discuss for classification are:

- ◇ Atlas based segmentation: This approach involves construction of an atlas or look up table corresponding to every organ/tissue recording its anatomical features. These methods perform segmentation and classification in one pass. The limitation however is building the atlas with exhaustive features per organ, for eventually all organs.

- ◇ Artificial Intelligence (AI) for segmentation and classification: Both supervised and unsupervised AI techniques have been used for this purpose. Supervised Artificial Neural Networks (ANNs) are self-organizing, adaptive and use training data to solve complex problems in real-time, courtesy parallel programming. In ANNs, creating a labeled training set and training is a bottleneck. Also, the performance is sensitive to training parameters and is adversely affected by the presence of noise. Unsupervised methods include clustering that overcome the bottleneck of training but bring with them another set of limitations. These include setting algorithm initiation parameters and stopping conditions and that the algorithms are prone to convergence at local minima.

A large number of image classification studies are carried out at the whole image level, that is by extracting features per image ([69], [35] and [56]), there are special cases where pixel classification is preferred.

Online et al. [59] adopt a pixel classification technique for skin segmentation. In face or gesture detection problems, detecting skin in images reduces the search space for objects of interest, such as faces or hands. The authors use a color pixel classification approach. They compare the various color spaces including the below channels and their combinations.

- ◇ RGB: All colors in the image are specified by three primary components: red, blue and green channels.

- ◇ HSV: Colors are described by the hue, saturation and intensity value channels. This color space is similar to HIS, HLS and HCL.

- ◇ YCbCr: Colors are specified by the luminance and chrominance channels.

The performance of four classifiers with pixel features extracted in each of the color spaces is also compared. The classifiers tested are Linear Decision Boundary, Naive Bayes, Gaussian and Multi-layer Perceptron. The study finds that classifier performance using color spaces with only chrominance channels are weaker than when all channels (RGB, HSV or YCbCr) are used. Further, Naive Bayes and Multi-layer perceptron classiffiers are seen to consistently perform better than the others. Extensions have also been proposed to color spaces. Macaire et al. [57] propose the use of a new hybrid color space that has *d* dimensions unlike the conventional three dimensional color spaces.

Deformable models are curves or surfaces that deform under the influence of internal and external image forces to demarcate object boundary hence segmenting the image. These models work on an energy minimization problem over the image based on edge or region based features. They involve no prior training step. Huang and Tsechpenakis [42] describe a new deformable model called *"Metamorphs"* that integrates both edge and region based features in the energy function together. With the edge energy term E_E and the region energy term E_R and a constant k, the Metamorphs energy function, E, is defined as

$$E = E_E + kE_R$$

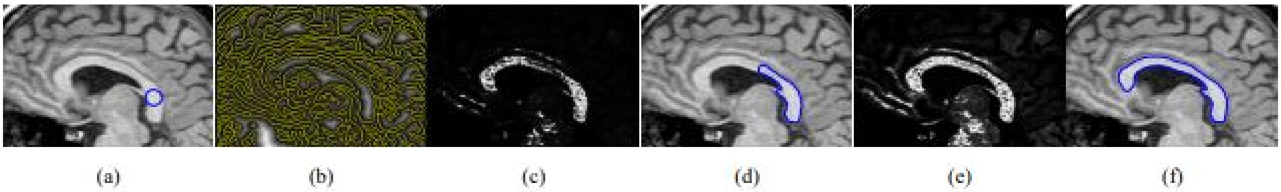


Figure 6.4: The image at various steps in the Metamorphs segmentation model. a) Original Image a brain structure with a initial selected region of Interest. b) Edges detected using Canny's edge detection algorithm. c),e) Intensity likelihood based on the internal energy forces. d) Intermediate evolving model after 15 iterations. f) Converged model after 38 iterations. [42]

Figure 6.4 shows the iterative steps of the Metamorphs model and its convergence on a frame of a brain structure. The metamorphs models are not learning based and hence do not need long training times. However, the result of the model hugely depends on the initialization and the model is prone to convergence at local minima. Unlike the authors' work here, we adopt learning based methods in this thesis. This does require long training time and a large training set but once the model is trained, the predictions can be very close to real-time. Like Huang and Tsechpenakis [42], we also use edge and region based features together.

6.2.1 Pixel-based Machine Learning

Extensive work on both pixel-based and object-based classification has been done. Pixel-based classification requires a feature vector per pixel in the image for training. On the other hand, in object-based classification, the image is first segmented together forming multiple joint and disjoint chunks of similar pixels. The classifier is trained with feature vectors corresponding to these segments. It is widely speculated that the object-based classifiers outperform the pixel-based classifiers. However, the accuracy of object-based classifiers strongly depends on the quality of underlying segmentation algorithm [53]. With the advancements in computing technologies, pixel-based machine learning techniques are being experimented with. Suzuki [78] survey the various pixel/voxel-based machine learning (PML) used on medical images to aid diagnostics. These PML techniques can avoid errors caused by inaccurate segmentation or feature generation. Due to these reasons, the performance of PML is better than object-based learning in some cases. Another issue with using features per image is the loss of localization. When features are summarized per image, their spatial origin is lost. Hence, making object detection cumbersome. There are three classes of PML algorithms that have been implemented on a variety of medical images- neural filters, convolution neural networks (CNNs) and massive-training artificial neural networks (MTANNs)

Neural filters are a class of PML algorithms that have been used for image processing. A filter plays an important role for improving the sensitivity and specificity in CAD schemes. The application of neural filters include edge-preserving noise reduction [79], edge enhancement from noisy images [80] and supervised edge enhancing [81]. The neural filters imply a linear artificial neural network model. The input to the neurons are the values of all the pixels in the neighborhood of the *current* pixel. The output is the pixel value of the *current* pixel. The final output image is formed by combining all the individual pixel values. In the supervised scenario, the filter is trained with a *teaching* image corresponding to every train image. The value of the error function is back-propagated through the ANN to improve its output in every iteration. The performance of the neural filter is found to be superior to a conventional averaging filter (Figure 6.5).

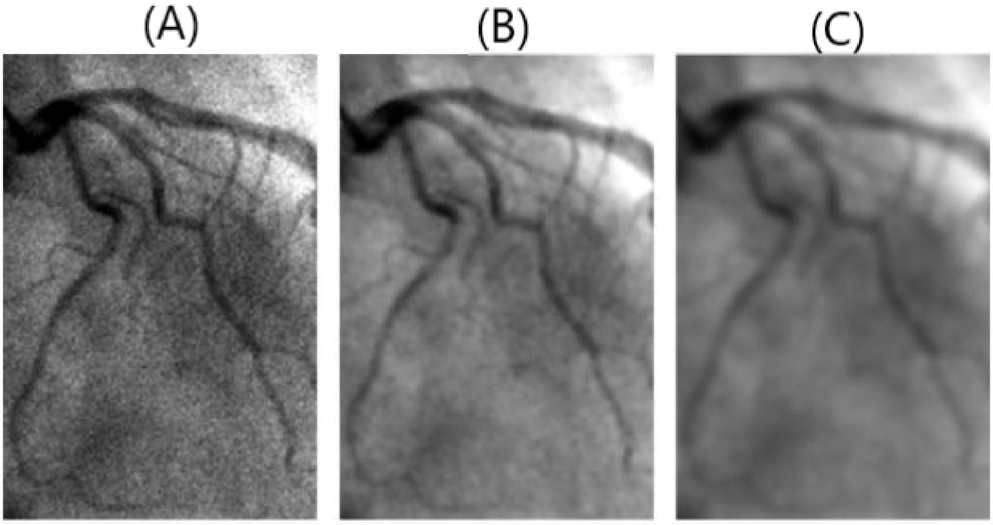


Figure 6.5: (A) A noisy angiogram input. Noise reduction by the (B) neural filter and (C) averaging filter. [78]

The second class of PML algorithms - convolution neural networks (CNNs) are widely used for image classification to support medical diagnostics. CNNs, responsible for major breakthroughs in image classification, are the core of most Computer Vision systems today. They consist of one input and several hidden layers. The layers of the CNN are connected with convolutions with a local kernel or filter (that are automatically learnt by the CNN in the training phase). The convolved values calculated, passed to an *activation function* and sent through the layers and finally exit the network at the output layer. The input to the CNN are all the pixel values of the image and the output is the class to which the input image belongs. However, in this basic implementation, all localization information is lost and object detection within the image is not possible. Shift-Invariant neural networks are an enhancement of CNN that produce images as output and not just class labels. Hence, making localization of objects within an image (eg. lesions) possible.

In Massive training artificial neural networks (MTANNs), The input to the MTANN are the pixels in a region of the image. The output is the center pixel of the input region (Figure6.6). To enrich training samples, multiple overlapping subregions are created in a region and training is repeated for all pixels in each subregion. Owing to the resulting *massive* number of training samples, this class of PML algorithms is appropriately named. The working unit of the MTANN can be any ML model of choice- feed-forward ANN, Support Vector Machines etc,. MTANNs have been successfully applied for image processing, segmentation, classification and also object detection. In [77] authors develop a supervised filter for the enhancement of actual lesions using MTANNs for detection of lung nodules in CT. This is also a good technique to reduce false positives. Suzuki et al. [84] use MTANNs to separate bones from soft tissues. Benign and Malignant lung tumors are distinguished by Suzuki et al. [82] using MTANNs.

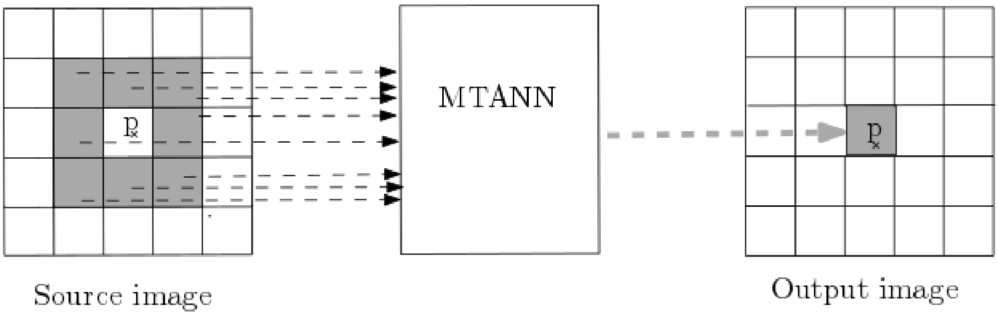


Figure 6.6: Explained input and output of the MTANN.

As stated earlier, image segmentation and classification are interlinked. A classifier implicitly segments an image and a segmentation implies a classification [1]. In this thesis we attempt to segment and classify tissues present in the images in one pass. We use a per-pixel approach for *classification* with features based on the pixel's RGB color space (brightness) but also additionally take the position of pixel into consideration. We use Canny's edge detection algorithm [26] to identify if a pixel belongs to an edge. Additionally unlike Sharma and Aggarwal [70], we use both these features together. This combination of features is found to enhance the segmentation results observed. We leverage supervised learning based AI techniques for classification. Further, we do not use single pixel values but also the pixel values of the locality or neighbourhood. This is explained in figure 6.7.

6.2.2 Video Processing Pipelines

Deep learning based Convolution Neural Networks (CNNs) have been known to perform extremely well for image and also speech applications and are currently used as the state-of-the-art performance benchmark as studied by Sermanet et al. [68],Sce [2] and Cires¸ancires¸an et al. [28]. They have been consistently performing better than traditional models that require handpicking of features from inputs. These models eliminate the feature generation, extraction and selection process that are key to deciding the performance of classical ML models. The building bocks of all deep learning models are layers of neurons with weights and activation functions. However, their arrangement can be modified to lead to different types of deep networks each recommended for specific use-cases. For instance, multilayer perceptron (MLP) consists of 2-3 hidden layers and are widely used for natural language processing. CNNs contain one or more convolution layers that filter the inputs to the net layer and hence allows the network to be *deeper* with fewer parameters. As mentioned above, they are extensively used for image processing and classification but also for video analysis [54]. Recurrent neural networks (RNN) contains a directed cycle in the connections between neurons. This allows temporal information effecting the output. This makes RNN suitable for connected handwriting recognition and text classification as shown bt Lai et al. [47].

However, Ye et al. [92] show that their performance has not been up-to expectations when it comes to video analysis. This is credited to two major factors: the spatial-temporal nature of videos and the limited availability of annotated video data for training. Ye et al. [92] approach

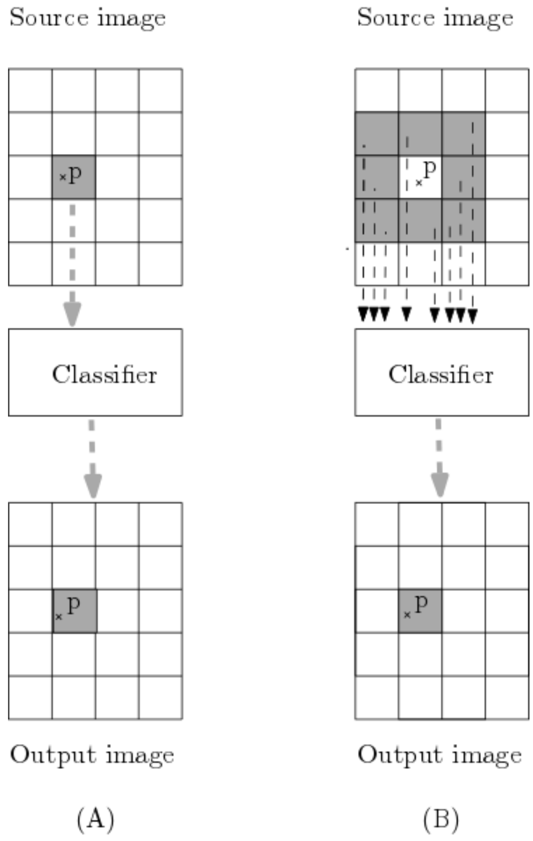


Figure 6.7: The two pixel-based learning approaches compared in this thesis.

this issue by handling video data training in two separate CNNs on spatial and temporal data respectively and fusing the two networks for the final output. The spatial data is the a stack of the static frames that make up the data and the temporal data is the a stack of the optical flow that shows the displacement vector between two frames. Karpathy et al. [45] evaluates the performance of this approach on the 200,000 videos in the Sports-1M dataset. In this work, we refrain from using RNN based architectures due to availability of limited data. Instead, we implement and test a CNN based network architecture - U-Net by Ronneberger et al. [65] which works well with small training set size too.

6.3 Object Detection

Identifying an object within an image, that is, locating its position, followed by predicting the class that it belongs to is called object detection. Object detection is modeled as a classification problem where multiple overlapping windows of all sizes from the input image are fed to the classifier. The problem of choosing the size of the sliding window is critical here and is tackled by resizing the image at multiple scales such that the chosen window size will completely contain an object in one of the scales chosen. This idea is illustrated in figure 6.8

Region based convolution neural networks (R-CNNs) are one of the algorithms that use classification as the underlying problem for object detection. An object proposal algorithm called ”Selective Search” reduces the number of bounding boxes per image using local features

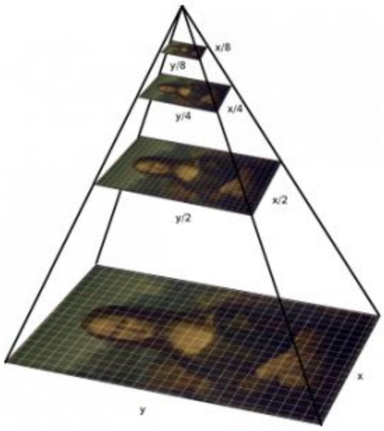


Figure 6.8: Resizing the image to multiple scales in order to fit the object in a single window in one of the resized images [15]

.

like color, texture, position etc,. The reduced bounding boxes are fed to a CNN classifier. In order to improve the speed, several modifications byGirshick [39], Ren et al. [63] and He et al. [41] have been proposed. Faster R-CNN by Ren et al. [63] replaces the Selective Search algorithm by a small convolution network called the Region Proposal Network.

Unlike conventional methods that handle detection as a classification problem, two algorithms handle it as a regression problem. These two most relevant state-of-the-art algorithms for object detection are *"You only look once"* (YOLO) and *"Single Shot Detector"* (SSD).

YOLO by Redmon et al. [62] divides the image into an *SxS* grid and creates *B* bounding boxes with different confidence levels and *C* class probabilities. This is depicted in figure 6.9. A threshold for the confidence can be set to limit the number of bounding boxes accordingly. YOLO sees an image just once and hence is very fast and used in real-time. Sindhu Ramachandran S. [73] detect nodules in real-time in the CT scans of the lung using a YOLO based deep net and shows a reduced false positive rate and high precision and sensitivity.

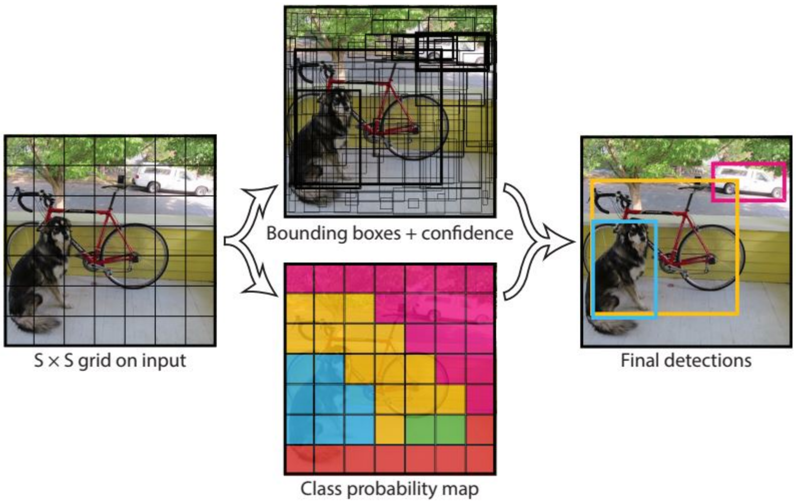


Figure 6.9: The steps involved in YOLO] [62]

SSD by Liu et al. [55] also runs an image through a CNN only once. The CNN generates a *feature map*. Small convolution networks are run on this feature map to predict bounding boxes and classification probability. SSD achieves a good balance of speed and accuracy. It can also be used real-time.

In this thesis, we implement a CNN based network called U-Net by Ronneberger et al. [65] that has gained popularity due to its high performance in the domain of medical image processing. The results are compared with that obtained using other more traditional pixel-based machine learning models.

7
Conclusion

Hepatocellular carcinoma (HCC), the fifth most common cancer worldwide, has led to 782,000 deaths all over the world as of 2012. It is the third highest cause for cancer related deaths. Over the years, research has led to several treatments for the cancer including resection, liver transplantation, radioembolization and chemoembolization. Today, HCC has stepped down from being an almost universal death sentence to a cancer that may be prevented and treated and cured if detected at an early stage [25]. With the advancements in image processing and digital systems several medical imaging techniques have risen to aid the early detection of HCC. Sonography or ultrasound imaging is an effective, portable and real-time imaging technique heavily used for the diagnosis of several liver malignancies. Furthermore, enhancement agents are introduced in the patients blood flow to additionally enhance the visibility of tumors. However, the diagnosis of HCC remains a challenge due to factors like short duration of arterial phase where malignancies can be spotted and individual patient history. We propose the use of artificial intelligence to assist physicians in the diagnosis of the disease.

In a nutshell, our system does the following. CEUS scan (video) data collected are converted to images. The images are labelled based on information we received from physicians. Image pre-processing is done to remove noise. The pre-processed images are used to train both traditional machine learning and deep neural network based classifiers. The predictions of the classifiers are visualized and evaluated by relevant performance metrics. The system is trained with data from multiple patients and hence overtime can learn to aid physicians in their diagnosis.

The machine learning algorithms experimented with include Naive Bayes, Hoeffding Tree, J48 Decision Tree, Random Forest, Sequential Minimal Optimization and k- Nearest Neighbours. The performance of each of the models was thoroughly analysed and it was found that Sequential Minimal Optimization and k- Nearest Neighbours show the best performance with an AUC of 0.8 and 0.77 respectively. The weighted precision and weighted recall of the SMO classifier are 0.65 and 0.62 respectively. On the other hand, the same for the k-NN classifier 0.76 and 0.59 respectively. The deep convolution network architecture called U-Net was also trained on the available CEUS data. The model performed very well with and intersection over union (IOU) of 95.89%.

Therefore, it is seen that neural network based architectures do indeed perform better than traditional machine learning algorithms in the current problem of tissue recognition in contrast enhanced ultrasound scans. However, they require longer training time and dedicated infrastructure.

Bibliography

[1] Oxford University- Information Engineering, Image Segmentation and Classification Lecture 5 Professor Michael Brady FRS FREng Hilary Term 2005. URL http://www.robots.ox.ac.uk/{~}jmb/lectures/InformaticsLecture5.pdf.

[2] Learning Hierarchical Features for Scene Labeling. URL http://yann.lecun.com/exdb/publis/pdf/farabet-pami-13.pdf.

[3] T.f.cootes, c.j.taylor. a mixture model for representing shape variation. proc 8th bmvc (vol.1) (ed. a.f.clark) bmva press, pp.110-119. 1997.

[4] Annotation Tool. https://github.com/maverickjoy/bounding-box-annotation-tool.

[5] Dicom standard- data structures and encoding. . URL http://dicom.nema.org/medical/dicom/current/output/pdf/part05.pdf.

[6] Dicom standard- concepts. . URL https://www.dicomstandard.org/concepts/.

[7] Dicom standard- data structure and encoding. . URL dicom.nema.org/medical/dicom/current/output/pdf/part05.pdf.

[8] Dicom Viewer. http://www.phon.ox.ac.uk/jcoleman/old_SLP/Lecture_5/DTW_explanation.html,.

[9] Dicom Viewer. http://trac.research.cc.gatech.edu/GART/browser/GART/weka/edu/gatech/gart/ml/weka/DTW.java?rev=9,.

[10] Mayo clinic. URL https://www.mayoclinic.org/.

[11] International contrast ultrasound society. URL http://www.icus-society.org.

[12] Karl theo dussik. URL http://www.ob-ultrasound.net/dussikbio.html.

[13] Dicom Viewer. http://www.microdicom.com/.

[14] Color spaces. URL http://www.opticallimits.com/colorimetric-systems-and-color-models.

[15] Guide to object detection using deep learning: Faster r-cnn,yolo,ssd. URL http://cv-tricks.com/object-detection/faster-r-cnn-yolo-ssd/.

[16] Ultrasound physics. URL http://www.vaultrasound.com/educational-resources/ultrasound-physics.

[17] How ultrasound works. URL https://www.physics.utoronto.ca/~jharlow/teaching/phy138_0708/lec04/ultrasoundx.htm.

[18] A Survey on Lossless Compression for Medical Images. *International Journal of Computer Applications*, 31-No.8, 2011. URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.259.1082{&}rep=rep1{&}type=pdf.

[19] S. Adhikari, E. Situ-LaCasse, J. Acuna, L. Friedman, E. Tay, J. Tsung, and M. Blaivas. 340 Accuracy of Point-of-Care Ultrasound for the Diagnosis of Scrotal Pathology in the Emergency Department. *Annals of Emergency Medicine*, 70(4):S135, oct 2017. ISSN 01960644. doi: 10.1016/j.annemergmed.2017.07.411. URL http://linkinghub.elsevier.com/retrieve/pii/S0196064417313112.

[20] T Albrecht, E Leen, R Lencioni, M Blomley, D Lindsell, L Bolondi, L Thorelius, and K Jäger. Guidelines for the Use of Contrast Agents in Ultrasound. *Ultraschall in Med*, 2004. ISSN 01724614. doi: 10.1055/s-2004-813245.

[21] Thomas Albrecht, Martin Blomley, Luigi Bolondi, Michel Claudon, J-M Correas, David Cosgrove, Lucas Greiner, Kurt Jäger, Nico De Jong, Eddie Leen, et al. Guidelines for the use of contrast agents in ultrasound-january 2004. *Ultraschall in der Medizin-European Journal of Ultrasound*, 25(04):249–256, 2004.

[22] Irene Bargellini, Valentina Battaglia, Elena Bozzi, Dario Luca Lauretti, Giulia Lorenzoni, and Carlo Bartolozzi. Radiological diagnosis of hepatocellular carcinoma. *Journal of Hepato-*

cellular Carcinoma, 1:137, sep 2014. ISSN 2253-5969. doi: 10.2147/JHC.S44379. URL http://www.ncbi.nlm.nih.gov/pubmed/27508183http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4918274http://www.dovepress.com/radiological-diagnosis-of-hepatocellular-carcinoma-peer-reviewed-article-JHC.

[23] Craig M Bennett, George L Wolford, and Michael B Miller. The principled control of false positives in neuroimaging. *Social cognitive and affective neuroscience*, 4(4):417–422, 2009.

[24] Marie-Odile Bernier, Neige Journy, Hélène Baysson, Sophie Jacob, and Dominique Laurier. Potential cancer risk associated with CT scans: Review of epidemiological studies and ongoing studies. *Progress in Nuclear Energy*, 84:116–119, sep 2015. ISSN 0149-1970. doi: 10.1016/J.PNUCENE.2014.07.011. URL https://www.sciencedirect.com/science/article/pii/S0149197014001929.

[25] Jordi Bruix, Morris Sherman, and American Association for the Study of Liver Diseases. Management of hepatocellular carcinoma: an update. *Hepatology (Baltimore, Md.)*, 53(3):1020–2, mar 2011. ISSN 1527-3350. doi: 10.1002/hep.24199. URL http://www.ncbi.nlm.nih.gov/pubmed/21374666http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3084991.

[26] John Canny. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6):679–698, 1986.

[27] Yong Eun Chung and Ki Whang Kim. Contrast-enhanced ultrasonography: advance and current status in abdominal imaging. *Ultrasonography (Seoul, Korea)*, 34(1):3–18, jan 2015. ISSN 2288-5919. doi: 10.14366/usg.14034. URL http://www.ncbi.nlm.nih.gov/pubmed/25342120http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4282229.

[28] Dan C Cires¸ancires¸an, Alessandro Giusti, Luca M Gambardella, and J Urgen Schmidhuber. Deep Neural Networks Segment Neuronal Membranes in Electron Microscopy Images. URL http://www.idsia.ch/.

[29] Michel Claudon, Christoph F Dietrich, Byung Ihn Choi, David O Cosgrove, Masatoshi Kudo, Christian P Nolsøe, Fabio Piscaglia, Stephanie R Wilson, Richard G Barr, Maria C Chammas, Hans-Peter Weskott, and Hui-Xiong Xu. GUIDELINES AND GOOD CLINICAL PRACTICE RECOMMENDATIONS FOR CONTRAST ENHANCED ULTRASOUND (CEUS) IN THE LIVER – UPDATE 2012 A WFUMB-EFSUMB INITIATIVE IN COOPERATION WITH REPRESENTATIVES OF AFSUMB, AIUM, ASUM, FLAUS AND ICUS. *Ultrasound in Medicine & Biology*, 39(2):187–210, 2013. doi: 10.1016/j.ultrasmedbio.2012.09.002. URL http://dx.doi.org/10.1016/j.ultrasmedbio.2012.09.002.

[30] Ge Cong and Bahram Parvin. Model-based segmentation of nuclei. *Pattern recognition*, 33(8):1383–1393, 2000.

[31] Gabriella Csurka, Christopher R Dance, Lixin Fan, Jutta Willamowski, and Cédric Bray. Visual Categorization with Bags of Keypoints. URL https://www.cs.cmu.edu/{~}efros/courses/LBMV07/Papers/csurka-eccv-04.pdf.

[32] C. F. Dietrich, M. A. Averkiou, J. M. Correas, N. Lassau, E. Leen, and F. Piscaglia. An EFSUMB introduction into dynamic contrast-enhanced ultrasound (DCE-US) for quantification of tumour perfusion. *Ultraschall in der Medizin*, 33(4):344–351, 2012. ISSN 01724614. doi: 10.1055/s-0032-1313026.

[33] Kunio Doi. Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. *Computerized Medical Imaging and Graphics 31 (2007) 198–211*. URL https://ac.els-cdn.com/S0895611107000262/1-s2.0-S0895611107000262-main.pdf?{_}tid=f6cfe35f-7e28-459d-8e67-a4459285b18e{&}acdnat=1529410720{_}a6dc4c5d01ce9743d04c72c809bb170a.

[34] Jacques Ferlay, Hai-Rim Shin, Freddie Bray, David Forman, Colin Mathers, and Donald Maxwell Parkin. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *International Journal of Cancer*, 127 (12):2893–2917, dec 2010. ISSN 00207136. doi: 10.1002/ijc.25516. URL http://doi.wiley.com/10.1002/ijc.25516.

[35] Steven Finette, Alan Bleier, ”", and William Swindell. BREAST TISSUE CLASSIFICATION USING DIAGNOSTIC ULTRASOUND AND PATTERN RECOGNITION TECHNIQUES: I. METHODS OF PATTERN RECOGNITION. *JLTRASONIC IMAGING*, 5: 55–70, 1983. URL https://s3.amazonaws.com/academia.edu.documents/46025521/0161-7346{_}2883{_}2990101-320160528-5403-1809wve.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A{&}Expires=1530820800{&}Signature=yzI8iij7FcU9dq{%}2F0Cz63JcFcZKY{%}3D{&}response-content-disposition=

inline{%}3Bfilename{%}3DBreast{_}tissue{_}classification{_}using{_}diagn.pdf.

[36] N. Frulio and H. Trillaud. Ultrasound elastography in liver. *Diagnostic and Interventional Imaging*, 94 (5):515–534, may 2013. ISSN 2211-5684. doi: 10.1016/J.DIII.2013.02.005. URL https://www.sciencedirect.com/science/article/pii/S2211568413000405.

[37] A Gelet, J Y Chapelon, T J Margonari, ! Y Theillere, F Gorry, ! D Cathignol, and E Blanct. Prostatic Tissue Destruction by High-Intensity Focused Ultrasound: Experimentation on Canine Prostate1. *JOURNAL OF ENDOUROLOGY*, 7(3), 1993. URL https://www.liebertpub.com/doi/pdf/10.1089/end.1993.7.249.

[38] Soumya Ghose, Arnau Oliver, Jhimli Mitra, Robert Martí, Xavier Lladó, Jordi Freixenet, Désiré Sidibé, Joan C Vilanova, Josep Comet, and Fabrice Meriaudeau. A supervised learning framework of statistical shape and probability priors for automatic prostate segmentation in ultrasound images. *Medical Image Analysis*, 17:587–600, 2013. doi: 10.1016/j.media.2013.04.001. URL https://www.medicalimageanalysisjournal.com/article/S1361-8415(13)00045-5/pdf.

[39] Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015.

[40] Luis Guerra, Laura M. McGarry, Víctor Robles, Concha Bielza, Pedro Larrañaga, and Rafael Yuste. Comparison between supervised and unsupervised classifications of neuronal cell types: A case study. *Developmental Neurobiology*, 71(1):71–82, jan 2011. ISSN 19328451. doi: 10.1002/dneu.20809. URL http://doi.wiley.com/10.1002/dneu.20809.

[41] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, pages 2980–2988. IEEE, 2017.

[42] Xiaolei Huang and Gavriil Tsechpenakis. Medical Image Segmentation. URL https://pdfs.semanticscholar.org/bdf3/accd867e283b5ca2d3e0371e773837744dc3.pdf.

[43] Hyun-Jung Jang, Tae Kyoung Kim, Peter N. Burns, and Stephanie R. Wilson. Enhancement Patterns of Hepatocellular Carcinoma at Contrast-enhanced US: Comparison with Histologic Differentiation. *Radiology*, 244(3):898–906, sep 2007. ISSN 0033-8419. doi: 10.1148/radiol.2443061520. URL http://pubs.rsna.org/doi/10.1148/radiol.2443061520.

[44] Jia Deng, Wei Dong, R. Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, (May 2014):248–255, 2009. ISSN 1063-6919. doi: 10.1109/CVPRW.2009.5206848. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5206848.

[45] Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Li Fei-Fei. Large-scale Video Classification with Convolutional Neural Networks. URL http://cs.stanford.edu/people/karpathy/deepvideo.

[46] Diederik P Kingma and Jimmy Lei Ba. ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION. Technical report. URL https://arxiv.org/pdf/1412.6980.pdf.

[47] Siwei Lai, Liheng Xu, Kang Liu, and Jun Zhao. Recurrent convolutional neural networks for text classification. 2015. URL https://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9745/9552.

[48] R. A. Lerski, M. J. Smith, P. Morley, E. Barnett, P. R. Mills, G. Watkinson, and R. N. M. MacSween. Discriminant Analysis of Ultrasonic Texture Data in Diffuse Alcoholic Liver Disease: 1. Fatty Liver and Cirrhosis. *Ultrasonic Imaging*, 3(2):164–172, apr 1981. ISSN 0161-7346. doi: 10.1177/016173468100300203. URL http://journals.sagepub.com/doi/10.1177/016173468100300203.

[49] R.A. Lerski, E. Barnett, P. Morley, P.R. Mills, G. Watkinson, and R.N.M. MacSween. Computer analysis of ultrasonic signals in diffuse liver disease. *Ultrasound in Medicine & Biology*, 5(4):341–343, jan 1979. ISSN 0301-5629. doi: 10.1016/0301-5629(79)90004-8. URL https://www.sciencedirect.com/science/article/pii/0301562979900048.

[50] Gang Li, Tianming Liu, J Nie, L Guo, J Chen, J Zhu, Weiming Xia, A Mara, S Holley, and STC Wong. Segmentation of touching cell nuclei using gradient flow tracking. *Journal of Microscopy*, 231(1):47–58, 2008.

[51] Yuanqing Lin, Fengjun Lv, Shenghuo Zhu, Ming Yang, Timothee Cour, Kai Yu, Liangliang Cao, Thomas Huang, and Beckman Institute. Large-scale Image Classification: Fast Feature Extraction and SVM Training. URL http://rogerioferis.com/VisualRecognitionAndSearch2014/material/papers/SuperVectorCVPR2011.pdf.

[52] Tony Lindeberg. Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention. *International Journal of Computer Vision*, 11(3):283–318, 1993.

[53] Desheng Liu and Fan Xia. Assessing object-based classification: advantages and limitations. *Remote Sensing Letters*, 1(4):187–194, dec 2010. ISSN 2150-704X. doi: 10.1080/01431161003743173. URL `http://www.tandfonline.com/doi/abs/10.1080/01431161003743173`.

[54] Li Liu and Ling Shao. Learning discriminative representations from RGB-D video data. *IJCAI International Joint Conference on Artificial Intelligence*, pages 1493–1500, 2013. ISSN 10450823.

[55] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *European conference on computer vision*, pages 21–37. Springer, 2016.

[56] David G Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004. URL `https://link.springer.com/content/pdf/10.1023/B:VISI.0000029664.99615.94.pdf`.

[57] Ludovic Macaire, Vincent Ultré, and Jack-Gérard Postaire. Determination of compatibility coefficients for color edge detection by relaxation Project with industry View project Motion analysis in image sequence View project. 1996. doi: 10.1109/ICIP.1996.561019. URL `https://www.researchgate.net/publication/224185285`.

[58] Tom M Mitchell et al. Machine learning. wcb, 1997.

[59] Research Online, Son Lam Phung, A Bouzerdoum, D Chai, Abdesselam Bouzerdoum, Sr Member, and Douglas Chai. Skin segmentation using color pixel classification: analysis and comparison Skin segmentation using color pixel classification: analysis and comparison Skin Segmentation Using Color Pixel Classification: Analysis and Comparison. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(1): 148–154, 2005. URL `http://ro.uow.edu.au/infopapers/256`.

[60] Kim Plunkett and Jeffrey L Elman. *Exercises in rethinking innateness: A handbook for connectionist simulations*. Mit Press, 1997.

[61] Syed A Quadri, Muhammad Waqas, Inamullah Khan, Muhammad Adnan Khan, Sajid S Suriya, Mudassir Farooqui, and Brian Fiani. Historical Remarks on HIFU High-intensity focused ultrasound: past, present, and future in neurosurgery. doi: 10.3171/2017.11.FOCUS17610. URL `https://thejns.org/doi/abs/10.3171/2017.11.FOCUS17610`.

[62] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You Only Look Once: Unified, Real-Time Object Detection. URL `http://pjreddie.com/yolo/`.

[63] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28*, pages 91–99. Curran Associates, Inc., 2015. URL `http://papers.nips.cc/paper/5638-faster-r-cnn-towards-real-time-object-detection-with-region-proposal-networks.pdf`.

[64] Jos BTM Roerdink and Arnold Meijster. The watershed transform: Definitions, algorithms and parallelization strategies. *Fundamenta informaticae*, 41(1, 2):187–228, 2000.

[65] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. may 2015. URL `http://arxiv.org/abs/1505.04597`.

[66] Iza Sazanita Isa, Siti Noraini Sulaiman, Muzaimi Mustapha, and Sailudin Darus. ScienceDirect Evaluating Denoising Performances of Fundamental Filters for T2-Weighted MRI Images-review under responsibility of KES International. *Procedia Computer Science*, 60:760–768, 2015. doi: 10.1016/j.procs.2015.08.231. URL `www.sciencedirect.com`.

[67] G. Schaefer, R. Starosolski, and Shao Ying Zhu. An evaluation of lossless compression algorithms for medical infrared images. In *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, pages 1673–1676. IEEE, 2005. ISBN 0-7803-8741-4. doi: 10.1109/IEMBS.2005.1616764. URL `http://ieeexplore.ieee.org/document/1616764/`.

[68] Pierre Sermanet, David Eigen, Xiang Zhang, Michael Mathieu, Rob Fergus, and Yann LeCun. OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks. dec 2013. URL `http://arxiv.org/abs/1312.6229`.

[69] Lior Shamir, John D. Delaney, Nikita Orlov, D. Mark Eckley, and Ilya G. Goldberg. Pattern Recognition Software and Techniques for Biological Image Analysis. *PLoS Computational Biology*, 6(11):e1000974, nov 2010. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1000974. URL `http://dx.plos.org/10.`

1371/journal.pcbi.1000974.

[70] Neeraj Sharma and Lalit M Aggarwal. Automated medical image segmentation techniques. *Journal of medical physics*, 35(1):3–14, jan 2010. ISSN 1998-3913. doi: 10.4103/0971-6203.58777. URL http://www.ncbi.nlm.nih.gov/pubmed/20177565http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2825001.

[71] Xuegong Shi, Roy W Martin, Daniel Rouseff, Shahram Vaezv ', and Lawrence A Crum+˜. Detection of High-Intensity Focused Ultrasound Liver Lesions Using Dynamic Elastometry. *ULTRASONIC IMAGING*, 21:107–126, 1999. URL http://journals.sagepub.com/doi/pdf/10.1177/016173469902100203.

[72] Junji Shiraishi, Qiang Li, Daniel Appelbaum, Kunio Doi, and Carl E Ravin. Computer-Aided Diagnosis and Artificial Intelligence in Clinical Imaging. *YSNUC*, 41: 449–462, 2011. doi: 10.1053/j.semnuclmed.2011.06.004. URL https://ac.els-cdn.com/S0001299811000742/1-s2.0-S0001299811000742-main.pdf?{_}tid=895f4f2e-4191-455f-98a5-a5426c9a2813{&}acdnat=1529437098{_}b209fbccf836bdc3e5d6df0a85e378db.

[73] Shibon Skaria Varun V. V. Sindhu Ramachandran S., Jose George. Using yolo based deep learning network for real time detection and localization of lung nodules from low dose ct scans. *Proc.SPIE*, 10575:10575 – 10575 – 9, 2018. doi: 10.1117/12.2293699. URL https://doi.org/10.1117/12.2293699.

[74] Josef Sivic and Andrew Zisserman. Video Google: A Text Retrieval Approach to Object Matching in Videos. 2003. URL http://www.robots.ox.ac.uk/{~}vgg/publications/papers/sivic03.pdf.

[75] Nitish Srivastava, Elman Mansimov, and Ruslan Salakhutdinov. Unsupervised Learning of Video Representations using LSTMs. URL http://proceedings.mlr.press/v37/srivastava15.pdf.

[76] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, and Ruslan Salakhutdinov. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. 15:1929–1958, 2014. URL http://www.cs.toronto.edu/{~}rsalakhu/papers/srivastava14a.pdf.

[77] Kenji Suzuki. A supervised 'lesion-enhancement' filter by use of a massive-training artificial neural network (MTANN) in computer-aided diagnosis (CAD). *Physics in Medicine and Biology*, 54(18):S31–S45, sep 2009. ISSN 0031-9155. doi: 10.1088/0031-9155/54/18/S03. URL http://stacks.iop.org/0031-9155/54/i=18/a=S03?key=crossref.5217392946fc8c1c9e56019916c71de8.

[78] Kenji Suzuki. Pixel-Based Machine Learning in Medical Imaging. *International Journal of Biomedical Imaging*, 2012:1–18, feb 2012. ISSN 1687-4188. doi: 10.1155/2012/792079. URL http://www.hindawi.com/journals/ijbi/2012/792079/.

[79] Kenji Suzuki, Isao Horiba, Noboru Sugie, and Michio Nanki. Neural filter with selection of input features and its application to image quality improvement of medical image sequences. *IEICE TRANSACTIONS on Information and Systems*, 85(10):1710–1718, 2002.

[80] Kenji Suzuki, Isao Horiba, and Noboru Sugie. Neural edge enhancer for supervised edge enhancement from noisy images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1582–1596, 2003.

[81] Kenji Suzuki, Isao Horiba, Noboru Sugie, and Michio Nanki. Extraction of left ventricular contours from left ventriculograms by means of a neural edge detector. *IEEE Transactions on Medical Imaging*, 23(3): 330–339, 2004.

[82] Kenji Suzuki, Feng Li, Shusuke Sone, and Kunio Doi. Computer-aided diagnostic scheme for distinction between benign and malignant nodules in thoracic low-dose ct by use of massive training artificial neural network. *IEEE Transactions on Medical Imaging*, 24(9):1138–1150, 2005.

[83] Kenji Suzuki, Junji Shiraishi, Hiroyuki Abe, Heber MacMahon, and Kunio Doi. False-positive reduction in computer-aided diagnostic scheme for detecting nodules in chest radiographs by means of massive training artificial neural network. *Academic radiology*, 12(2):191–201, feb 2005. ISSN 1076-6332. doi: 10.1016/j.acra.2004.11.017. URL http://www.ncbi.nlm.nih.gov/pubmed/15721596.

[84] Kenji Suzuki, Hiroyuki Abe, Heber MacMahon, and Kunio Doi. Image-processing technique for suppressing ribs in chest radiographs by means of massive training artificial neural network (mtann). *IEEE Transactions on medical imaging*, 25(4):406–416, 2006.

[85] M Tang, H Mulvana, T Gauthier, and A K P Lim. Quantitative contrast-enhanced ultrasound imaging : a review of sources of variability. (May):520–539, 2011.

[86] R Van Tiggelen and E Pouders. Ultrasound and computed tomography: spin-offs of the world wars. *JBR-BTR : organe de la Societe royale belge de radiologie (SRBR) = orgaan van de Koninklijke Belgische Vereniging*

voor Radiologie (KBVR), 86(4):235–41. ISSN 0302-7430. URL http://www.ncbi.nlm.nih.gov/pubmed/14527067.

[87] L Vincent and P Soille. Image-proc skeleton segmentation. *IEEE Trans Pattern Anal Mach Intell*, 13: 583–598, 1991.

[88] Joost Vromen and Brendan McCane. Red blood cell segmentation using guided contour tracing. 2006.

[89] Jiang Wang, Yi Yang, Junhua Mao, Zhiheng Huang, Chang Huang, and Wei Xu. CNN-RNN: A Unified Framework for Multi-label Image Classification. URL https://www.cv-foundation.org/openaccess/content{_}cvpr{_}2016/papers/Wang{_}CNN-RNN{_}A{_}Unified{_}CVPR{_}2016{_}paper.pdf.

[90] Jenna Wiens, John V Guttag, and Eric Horvitz. Patient Risk Stratification for Hospital-Associated C. diff as a Time-Series Classification Task. Technical report. URL http://erichorvitz.com/nips2012{_}CDiff{_}temporal.pdf.

[91] Joseph Woo. A short history of the development of ultrasound in obstetrics and gynecology. *History of Ultrasound in Obstetrics and Gynecology*, 3:1–25, 2002.

[92] Hao Ye, Zuxuan Wu, Rui-Wei Zhao, Xi Wang, Yu-Gang Jiang, and Xiangyang Xue. Evaluating Two-Stream CNN for Video Classification. 2015. doi: 10.1145/2671188.2749406. URL http://arxiv.org/abs/1504.01920{%}0Ahttp://dx.doi.org/10.1145/2671188.2749406.

[93] Hui Zhang, Jason E Fritts, and Sally A Goldman. Image segmentation evaluation: A survey of unsupervised methods. *computer vision and image understanding*, 110(2):260–280, 2008.