7-2019

Automatic Cancer Tissue Detection Using Multispectral Photoacoustic Imaging

Kamal Jnawali

kj5500@rit.edu

Automatic Cancer Tissue Detection Using Multispectral Photoacoustic Imaging

by

Kamal Jnawali

A dissertation submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in Center for Imaging Science

College of Science
Rochester Institute of Technology

July, 2019

Signature of the Author _____

Certified by _____
　　　　　　　PhD Program Director　　　　　　　　　　Date

PhD in Center for Imaging Science

ROCHESTER INSTITUTE OF TECHNOLOGY

ROCHESTER, NEW YORK

CERTIFICATE OF APPROVAL

Ph.D. DEGREE DISSERTATION

The Ph.D. Degree Dissertation of Kamal Jnawali
has been examined and approved by the
dissertation committee as satisfactory for the
dissertation required for the
Ph.D. degree in Center for Imaging Science

Dr. Navalgund Rao, Dissertation Advisor

Dr. Pengcheng Shi, Outside Chair

Dr. Vikram Dogra

Dr. Guoyu Lu

Date

ii

Automatic Cancer Tissue Detection Using Multispectral Photoacoustic Imaging

by

Kamal Jnawali

Abstract

Convolutional neural networks (CNNs) have become increasingly popular in recent years because of their ability to tackle complex learning problems such as object detection, and object localization. They are being used for a variety of tasks, such as tissue abnormalities detection and localization, with an accuracy that comes close to the level of human predictive performance in medical imaging. The success is primarily due to the ability of CNNs to extract the discriminant features at multiple levels of abstraction.

Photoacoustic (PA) imaging is a promising new modality that is gaining significant clinical potential. The availability of a large dataset of three dimensional PA images of ex-vivo human prostate and thyroid specimens has facilitated this current study aimed at evaluating the efficacy of CNN for cancer diagnosis. In PA imaging, a short pulse of near-infrared laser light is sent into the tissue, but the image is created by focusing the ultrasound waves that are photoacoustically generated due

to the absorption of light, thereby mapping the optical absorption in the tissue. By choosing multiple wavelengths of laser light, multispectral photoacoustic (MPA) images of the same tissue specimen can be obtained. The objective of this thesis is to implement deep learning architecture for cancer detection using the MPA image dataset.

In this study, we built and examined a fully automated deep learning framework that learns to detect and localize cancer regions in a given specimen entirely from its MPA image dataset. The dataset for this work consisted of samples with size ranging from $12 \times 45 \times 200$ pixels to $64 \times 64 \times 200$ pixels at five wavelengths namely, 760 nm, 800 nm, 850 nm, 930 nm, and 970 nm.

The proposed algorithms first extract features using convolutional kernels and then detect cancer tissue using the softmax function, the last layer of the network. The AUC was calculated to evaluate the performance of the cancer tissue detector with a very promising result. To the best of our knowledge, this is one of the first examples of the application of deep 3D CNN to a large cancer MPA dataset for the prostate and thyroid cancer detection.

While previous efforts using the same dataset involved decision making using mathematically extracted image features, this work demonstrates that this process can be automated without any significant loss in accuracy. Another major contribution of this work has been to demonstrate that both prostate and thyroid datasets can be combined to produce improved results for cancer diagnosis.

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my advisor Professor Dr. Navalgund Rao for his guidance, patience, and intellectual support throughout my Ph.D. research. I feel very fortunate to get a chance to work with him who offered invaluable research advice and always provided me opportunities to participate in collaborative research projects and conferences that allowed me to grow academically and intellectually. I am also very grateful to my Committee, Dr. Vikram Dogra, Dr. Guoyu Lu, and Dr. Pengcheng Shi for their time for providing helpful suggestions and feedbacks during my Ph.D. thesis preparation, without whom this accomplishment would not even have been possible.

I am very thankful to the Center for Imaging Science for all the support throughout the Ph.D. study. I would like to thank Center for Imaging Science, and the director of the center for imaging science Dr. David Messinger for providing me with a travel grant which allowed me to attend and present my works at Medical Imaging, SPIE, Defense + Commercial Sensing, SPIE, and International Conference on Semantic Computing, IEEE conferences. I would also like to extend my thanks to Elizabeth Lockwood for her countless assistance in my study arrangement during my Ph.D. study.

I would like to extend my sincere thanks to Dr. Mohammad R Arbabshirani, and Dr. Aalpen A. Patel at Geisinger Health System (GHS) for allowing me to work on the large scale radiological report and CT image datasets for brain hemorrhage detection.

I would like to thank Manoj Acharya, Dilip Aryal, Binyul Bajracharya, Pradeep

Abbreviation

PA	Photoacoustic
PAI	Photoacoustic Imaging
EM	Electromagnetic
2D	Two Dimensional
3D	Three Dimensional
MPA	Multispectral Photoacoustic
NIR	Near Infra-Red
X-ray CT	X-ray Computed Tomography
US	Ultrasound
PET	Positron Emission Tomography
SPECT	Single Photon Emission Computed Tomography
TRUS	Trans-Rectal Ultrasound
Nd:YAG	Neodymium-doped Yttrium Aluminum Aarnet
IRB	Institutional Review Board
CNN	Convolutional Neural Network
DCNN	Deep Convolutional Neural Network
ROC	Receiver Operating Characteristic
AUC	Area Under Curve

Contents

List of Figures

xxiii

List of Tables

Chapter 1

Introduction and Overview

Siegel et al. [88] reported 595,690 deaths out of 1,685,210 new cases of cancer in the United States in the year of 2016. 1,980 deaths out of 64,300 reported new cases of thyroid cancer, and 26,120 deaths out of 180,890 reported new cases of prostate cancer in the same year [88]. American Cancer Society recommends men older than 50 years for (prostate) cancer screening [28]. Screening is done with prostate-specific antigen and digital rectal examination, but both screenings suffer from low specificity (percentage of actual negatives that are correctly classified) and sensitivity (percentage of actual positives that are correctly classified) [28, 88].

Early detection and risk assessment play a significant role in successful treatments to improve the survival of patients [28, 69]. Ultrasound (US) imaging is widely used in cancer tissue [87] screening because of its relative safety, low cost, noninvasive nature, real-time display [67]. Transrectal ultrasound (TRUS) is a com-

mon US imaging for prostate cancer detection; however, there are types of cancers not visible to TRUS [28]. This is because the US has a relatively low contrast, which is based on the detection of mechanical properties of the tissue [45]. Incorrect diagnosis of the cancer tissues using US imaging could lead to delayed diagnosis and treatment [101] because of the low-resolution structural imaging due to the mechanical properties of tissue structure [28]. In addition to the US, magnetic resonance imaging (MRI) is another imaging for cancer tissue screening; however, it has not been widely used because of its high cost and slow imaging speed [89]. X-ray computed tomography (CT) offers high tissue penetration depth and excellent spatial resolution; however, it suffers from low sensitivity for cancer diagnosis [25]. Gonzalez et al. [25] also reported that X-ray radiation possesses risk of radiation associated with cancers [45]. Sinha [89] reported that CT has minimal application in cancer tissue screening. This is because of the risk of increased ionizing radiation exposure [45]. Positron emission tomography (PET) is capable of extracting functional information such as blood metabolism for cancer tissue detection [70]; however it suffers from low spatial resolution [45]. Single-photon emission computed tomography (SPECT) shares similar physics of imaging technique with PET, and hence, it also suffers from lower spatial resolution [89]. Currently, the primary medical imaging modality used for cancer diagnosis are US, CT, MRI, PET, and SPECT. Although most of these technologies are well established and widely used in practice, there are problems related to low sensitivity and specificity as in US/CT, radiation exposure as in CT, and higher cost as in MRI for cancer diagnosis [43].

Photoacoustic imaging (PAI) is a new medical imaging technique [104] that is currently making a transition from bench to bedside, both in terms of technology [104] and clinical applications [28]. PAI offers structural and functional imaging [81] in numerous medical imaging application areas such as blood vessels imaging [72, 72], cancer diagnosis [28, 72, 73], brain imaging [112] and is capable of high-resolution imaging [99]. PAI is based on the photoacoustic(PA) effect, which is a phenomenon of generating acoustic waves from an object illuminated by a pulsed laser light [104]. PAI exploits the PA effect to combine the strength of optical imaging [77]which is capable of producing high contrast imaging and ultrasound imaging which is capable of producing high resolution in deep tissue imaging [47, 77, 103]. PAI is safe due to its nonionizing radiation properties [104] greatly enhance its clinical practicability in the future, and it is also low-cost, which is an advantage over other imaging techniques like MRI or CT [34]. Hoelen et al. demonstrated that PAI could be implemented to image blood concentration around tumors [40] and is a key feature to detect cancer in the early stages of life [28, 110]. MRI or CT is not capable of detecting cancer tissue in the early stage and is capable of detecting cancer tissue when the diameter grows to 1 cm in size [61]. The PA imaging at the specific wavelength is capable of generating high contrast PA images of optically active corresponding chromophore as in figure (1.1) with high spatial resolution [21, 78]. Multiple high contrast PA images of the optically active chromophores such as oxyhemoglobin (high in oxygen content), deoxyhemoglobin (low in oxygen content), lipid, and water are generated by choosing multiple wavelengths to quantify the concentration of

the chromophores [23]. These chromophores' concentration makes the PA imaging system to image optical biomarker for cancer tissue detection [28, 30, 75, 90]. In multispectral photoacoustic (MPA) imaging, the large difference in light absorption coefficient between blood and other tissue constituents enables detection of tissue angiogenesis associated with rapid tumor growth in early stages [61]. Furthermore, there is strong evidence of the dependence of the frequency content of the PA signal with the size of the tumor [47, 91] (see chapter 3, section 3.2). During tissue pathology, the structural change of a tumor is expected, and hence, we hypothesized that frequency content would help to detect cancer [47, 91]. The use of the wavelength and frequency content while imaging a tissue specimen makes PAI is capable of structural and functional imaging. The PA image acquisition produces robust and less error-prone coregistered images capable of both structural and functional imaging compared to current imaging techniques such as MRI with US [3]. This is because the same ultrasound traducer can be used to record US and PA signals to generate a US image and PA image, where US and PA imaging modality are responsible for extracting structural and functional information of the chromophores respectively. The functional information such as concentrations of the chromophores and structural information such as the volumetric distribution of chromophore can be extracted with the use of multispectral (at multiple wavelengths) 3D (at multiple depths of a tissue specimen) PA imaging for cancer diagnosis [28, 109].

Even though any electromagnetic (EM) wave can generate the PA effects, there should be a careful choice of the range of EM for practical PAI [45]. The range
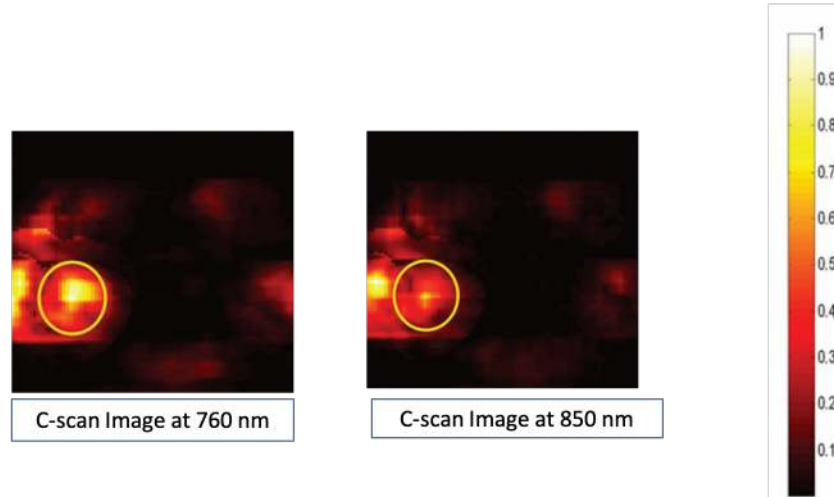
Figure 1.1: The figure shows the pixel intensity distribution of the encircled region of interest of the PA images taken at two wavelengths. The figure shows that the PA image is capable of detecting cancer because of the high concentration of the deoxyhemoglobin at 760 nm compared to the lower concentration of oxyhemoglobin at 850 nm. Reproduced with permission from [28].

of wavelength chosen for MPA imaging lies in the near-infrared region (NIR) for deep tissue imaging for two reasons. The first reason is that water is transparent within the window of the near-infrared region and is the significant content of the tissue. The second reason is that light suffers less from absorption and scattering in deep tissue imaging [11, 104]. In the case of the tissue imaging, four variables, namely deoxyhemoglobin, oxyhemoglobin, lipid, and water, were chosen to explain the variations and characteristics of malignant and nonmalignant regions [19, 20]. In this thesis, the five wavelengths from the NIR window, namely 760 nm, 850 nm, 930 nm, 970 nm, and 800 nm, were chosen. The first four wavelengths correspond to the peak of the absorption spectrum of deoxyhemoglobin, oxyhemoglobin, lipid, and water respectively, and the fifth wavelength corresponds to the equal of the absorption

spectrum of oxyhemoglobin and deoxyhemoglobin. The wavelengths were chosen in such a way that one variable is decreasing and the other is increasing to analyze the contribution of the individual chromophores for classification of malignant and normal region [20]. The choice of five wavelengths for cancer tissue detection was already proposed and evaluated in the literature [28, 75, 104]. The constituent and the concentration of the chromophores contain essential information related to a tissue abnormality [89]. The concentrations of deoxyhemoglobin and oxyhemoglobin provide information about angiogenesis for malignant tissue [91]. Dogra et al. [28] reported that malignant tumors are generally expected to be more oxygen-deficient than benign and healthy tissue. Figure (8.1) shows a brighter encircled region corresponds to be more oxygen-deficient (deoxyhemoglobin chromophore activated more at 760 nm) region than the normal tissue (oxyhemoglobin chromophore activated less at 850 nm). That confirms the presence of cancer in the given tissue specimen [28].

With the assumption that change in the tissue structure and size during tissue pathology, the frequency content of the PA signal is also a significant contributor to differentiate the malignant and normal tissue with three-dimensional (3D) PA imaging [89]. The chromophores concentration, and recorded PA frequency content at the given wavelength depend on tissue properties, and size and structure of chromophore [47, 89, 104] and hence the use of 3D PA (multiple C-scan across the entire depth of a tissue specimen) PA image at five wavelengths (chromophores' architectures and properties can be extracted with the use of absorption coefficient and is

highly dependent on the choice of wavelengths) is useful for cancer tissue diagnosis [28].
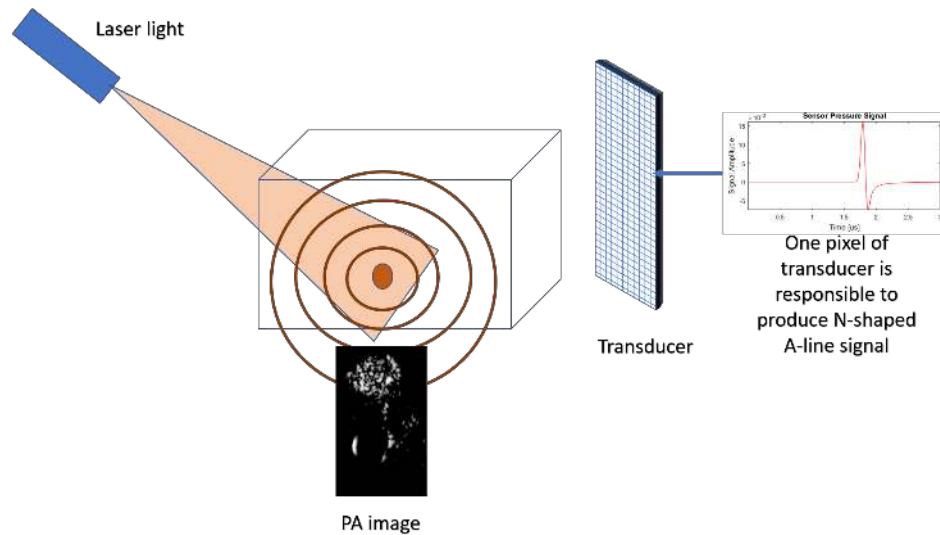
Figure 1.2: PA signal generation: The tissue was shone with short-pulsed laser light to generate PA signal, and was recorded by the US transducer.

In the PA effect, the tissue is generally irradiated by a short-pulsed laser [104]. When light is locally absorbed, heat is generated, which in turn produces an increased pressure. The increased pressure propagates as an ultrasound wave and generates a PA wave signal. The Nd: YAG lasers [111], in near-infrared (NIR) window, have been widely used as the excitation source in PA tissue imaging [102]. The pulse repetition rate of Nd: YAG lasers are typically low around 10 Hz [104]. Such lasers generate ten ns pulses with pulse energy in the range of tens of millijoules (mJ). This energy range falls within the American National Standards Institute (ANSI) limit [77] and hence is safe for cancer tissue imaging unlike other conven-

tional CT, PET and SPECT imaging [89]. The cost of PA imaging modality is very low compared to the other imaging modality [89] that gives a huge plus to the future for PA imaging [34]. Figure (1.2) shows the schematic of PA signal acquisition.

1.0.1 Outline

The contribution of this thesis is the development of an automated cancer tissue detector using the entire thyroid and prostate MPA image using deep learning algorithms. The cancer tissue detection technique is different from previous studies [28, 75, 91], which use pixel-based samples extracted from the encircled region of interest consisting of normal and malignant pixels and handcrafted features. The extraction of the encircled region in the PA image requires co-registration of PA image with histopathology slide and photograph image. The process of co-registration was very labor-intensive and time-consuming. The previous methods [28, 75] were trained, validated, and evaluated on the pixel-based samples extracted from the encircled region (figure (1.1)), not on the whole MPA image (see chapter 5). The co-registration of the MPA image with histopathology slide and photograph image is still required to extract the encircled region of interest during testing. These studies still require to provide the ground truth by the pathologist in advance during testing. This is because those methods require to extract a pixel-based sample from the encircled region of interest consisting of cancer and non-cancer region during test time. The previous studies are a tool to validating the cancer diagnosis proposed by the pathologist. Thus these studies still require manual work for the

detection of cancer during test time [28, 29, 75]. In this thesis, we made cancer tissue detection fully automated using an entire MPA image. We also implemented the grad-cam algorithm [84] for the automatic localization of the cancer tissue. To the best of our knowledge, this is one of the first applications of transfer learning using inception-resnet-v2 [96], and deep 3D convolutional neural network (CNN) for automatic cancer tissue detection and localization on the largest available cancer MPA image dataset.

The major contribution of this study is

- development of an automated cancer tissue feature extractor and detector using deep 2D/3D CNNs using the whole MPA image with the limited number of the dataset. This study circumvents the manual labor-intensive and time-consuming work of the previous studies [28, 29, 75] for the co-registration and extraction of handcrafted mathematical features.

- implementation of a grad-cam algorithm for an automatic localization of the cancer tissue region in a given MPA image.

1.1 Current Work: Application of Deep Learning

CNNs are increasingly popular in recent year because of their ability to tackle the complex learning problems such as object detection and object localization, close to the human level predictive performance. Recently, CNNs are introduced in medical image classification [31, 60, 95] and later to image segmentation [10, 24, 71], and

medical text classification [18, 49, 114] close to human-level predictive performance. The success is due to the ability of CNNs to extract the discriminant features at multiple levels of abstraction [54]. A convolutional layer in a CNN can detect certain local features of input images in all locations. A convolutional layer with k kernels detects k local features, where each filter has a set of shared weights and is called a convolutional kernel; the contribution of each kernel can be visualized with a feature map [56, 64, 117]. The convolutional layers in the CNN are often followed by the max-pooling layer, which reduces the feature map by selecting the maximum feature response. This is helpful for the transnational and small rotational invariance to some extent of an angle [15]. A neuron in the convolutional neural network has a very limited receptive field with shared weights. Therefore, fully connected layers are typically used after the convolutional layer. The choice of function in the final layer is problem specific. In this paper, a softmax function [50], and sigmoid function [51] were used for cancer tissue detection.

Recently, deep learning algorithms are becoming popular in the PAI domain [82] for PA image reconstruction [5, 37]. However, machine learning algorithms such as logistic function [28], two-layer neural network [89] and support vector machine (SVM) [63, 89] are still a popular choice in medical imaging applications such as cancer diagnosis. This thesis introduces the three deep learning algorithms to detect and localize the malignant region using the MPA image dataset [27].

The thesis work is divided into three parts, where the first part describes the implementation of the transfer learning [86] using inception-resnet-v2 [96] for the

thyroid cancer tissue detection using three-channel C-scan MPA images, the second part describes the implementation of the 3D CNN with seven layers for thyroid 105-slice C-scan MPA image dataset with higher detection accuracy than the transfer learning network. The third part describes the implementation of the 3D CNN with 11 layers on the mixture of thyroid and prostate 105-slice C-scan MPA image dataset with the best detection accuracy compared to the transfer learning and 3D CNN with seven layers. The third section also describes the implementation of the grad-cam algorithm for the localization of the cancer tissue using the deeper 3D CNN automatically using the MPA image.

1.1.1 Automatic Feature Extraction, Detection, Localization of Cancer Tissue Using MPA image dataset

In this thesis, we implemented the 2D transfer learning using inception-resnet-v2 [96] and deep 3D CNN with seven layers [51] were implemented for the cancer tissue detection using the thyroid MPA image dataset and the deep 3D CNN with 11 layers was implemented with the mixture of thyroid and prostate MPA dataset with the localization of cancer region in the test MPA image dataset (chapter 8). The same dataset was used in previous studies [28, 29, 75] for cancer tissue detection using machine learning methods with handcrafted features.

Transfer learning using inception-resnet-v2

We prepared the MPA image dataset for the transfer learning network using the first three wavelengths 760 nm, 800 nm, and 850 nm. The dimension of the PA dataset was ranging from $20 \times 64 \times 200$ pixels to $64 \times 64 \times 200$ pixels. The single-channel 2D C-scan image with maximum pixel intensity was extracted from the 200-channels 2D C-scan at each wavelength. The MPA image samples were prepared by concatenating three single-channel C-scan images at three wavelengths, namely 760 nm, 800nm, 850 nm. The spatial dimension of the MPA image was increased to 299×299 pixels using bilinear interpolation to fit into the input of the inception-resnet-v2 network. One sample of thyroid cancer tissue specimens and its PA is shown in figure [1.3]. The data structure used for the transfer learning technique is shown in figure (1.4)

Thyroid specimen   Histopathology slide   PA image cube
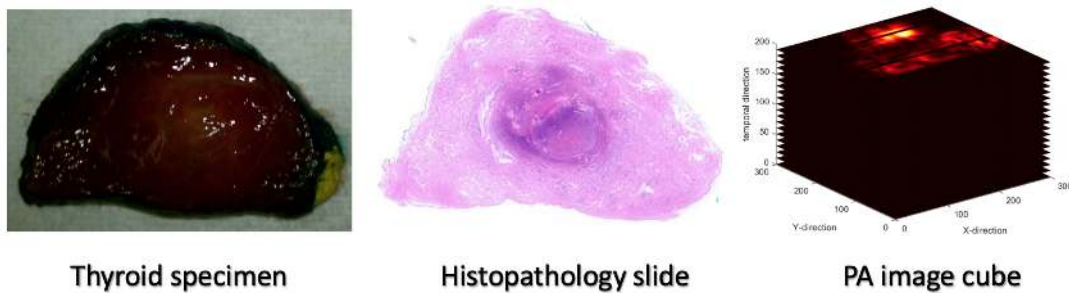
Figure 1.3: The figure shows the thyroid specimen with the metric scale, histopathological slide, and 3D PA image. The first two dimensions of the PA image cube corresponds to the spatial 2D C-scan image, and the third dimension corresponds to the A-line signals along the depth direction. One C-scan slice corresponded to the tissue of depth $4\mu m$. The spatial resolution of the PA image is 0.7×0.7 mm^2.
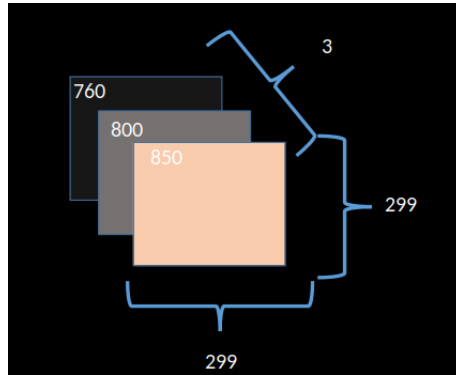
Figure 1.4: The input data structure of the inception-resnet-v2 consisted of 3-channel MPA image with $299 \times 299 \times 3$ pixels.

Tissue Specimen	Thyroid
Normal	40
Benign	46
Cancer	17
Total	103

Table 1.1: Normal, benign and malignant tissue distribution of thyroid MPA image dataset

Deep 3D Convolutional Neural Network

Deep 3D CNN was implemented to improve the predictive performance for cancer tissue detection than the transfer learning network. The structure of the input data of the transfer learning network was the concatenation of one-channel C-scan images at three wavelengths. The cancer tissue spreads in not only the 2D C-scan but also the depth direction. The transfer learning network was not capable of extracting the depth information for the spread of the cancer tissue in the depth direction.

The implementation of deep 3D CNN becomes useful when the depth information also becomes important [48]. In this thesis, to incorporate the depth and 2D C-scan details at once (volumetric information), we implemented deep 3D CNN for the cancer tissue detection with higher predictive performance.

The MPA dataset was collected in the imaging sciences lab at the University of Rochester. The PA images of each thyroid samples were generated with the short-pulsed laser at five different wavelengths: 760 nm, 800nm, 850 nm, 940nm, and 970 nm [28]. The set of wavelengths was chosen in order to extract the maximum information of the oxyhemoglobin and deoxyhemoglobin content in human tissue [104]. Those contents are prominent features for cancer tissue detection [75]. In this section, the samples were divided into noncancer and cancer while in the previous studies, the samples were divided into cancer, benign and normal tissue group [28, 29, 75, 90]. A radiologist provided the ground truth annotation for the MPA dataset. For this thesis, the MPA images were labeled as cancer if there was cancer in the given specimen otherwise labeled as the normal MPA image dataset which is different from the previous studies [28, 29, 75]. The size of the data cube at each wavelength was ranging from $20 \times 64 \times 200$ pixels to $64 \times 64 \times 200$ pixels. One sample of the prostate cancer tissue specimen and its PA image is shown in figure (1.5).

The 21-slices 2D C-scan image cube was taken at each wavelength. The spatial dimension of each PA sample was increased to 64×64 pixels with bilinear interpolation to make the size of all dataset uniform. The MPA sample images were prepared by concatenating five 21-slices 2D C-scan corresponding to 760 nm, 800 nm, 850
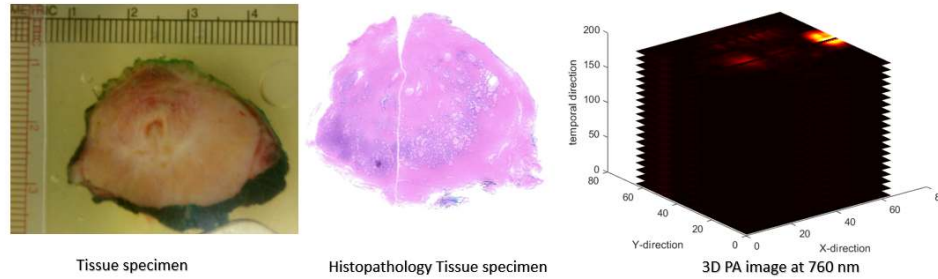
Figure 1.5: The figure shows the photograph, histopathology slide and 3D PA image (at 760 nm) of a prostate specimen respectively. The photograph was taken by a camera, the histopathological slide was prepared by the pathologist, and 3D PA image at 760 nm was prepared by the PA effect.

nm, 930 nm, and 970 nm, respectively, with dimensions equal to $64 \times 64 \times 105$ (1.6).

Deep 3D CNNs were implemented for two cases. The first case was the implementation of the deep 3D CNN with seven layers on the thyroid MPA image dataset. The second case was the implementation of deep 3D CNN with eleven layers on the mixture of prostate and thyroid MPA image dataset with the best performance. The latter case was able to detect the prostate and thyroid cancer at once with localization.

Tissue Specimen	Thyroid	Prostate	Total
Normal	91	17	108
Cancer	17	15	28
Total	108	28	136

Table 1.2: Normal and malignant tissue distribution

Automatic Localization of cancer tissue using MPA imaging

In this thesis, we implemented a grad-cam algorithm [84] to the proposed deep 3D CNN for automatic localization of the cancer tissue.
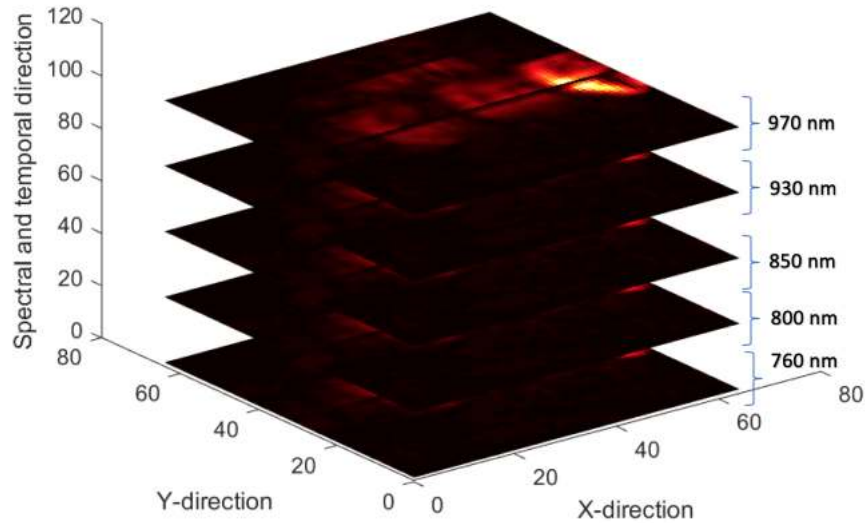
Figure 1.6: MPA image data structure with the dimension of $64 \times 64 \times 105$ pixels was proposed for the deep 3D architecture to incorporate the volumetric distribution of chromophores (frequency content of PA signal), and spectral signature (at five wavelengths) of the chromophores [28]. 21-slices C-scans images at five wavelengths were concatenated to make 105-channel C-scan image (21-slices C-scan $\times$ five wavelengths).

1.1.2   Architecture design and hyperparameter tuning

There are numerous choice to select number of CNN layers, number of fully connected layers, number of nodes in the fully connected layers and so on to design a typical optimal deep learning architecture for the classification task with the particular problem at hand [12]. Thus the search of the optimal set of hyperparameters makes the problem exponentially expensive in time [4]. For the efficient hyperparameter search, the architecture design for the cancer tissue detection was motivated

from the deep learning architecture implemented for the brain hemorrhage (see Appendix A).

1.2 Challenges faced and summary of logic behind the design of the network for the current dataset

In this thesis, deep learning algorithms were implemented for automatic cancer detection. In general, a good number of sample images are required to train deep neural networks to get the desired performance. The available dataset for this thesis was limited. However, there are some techniques such as transfer learning [66] that have proven to be useful for training in such situations. The first project of the thesis was to implement transfer learning using inception-resnet-v2 [96]. When data is very similar to the imagenet image [26], then the training of the network is only required on the last softmax layer of the inception-resnet-v2 [53]. But the medical datasets are in general different from the imagenet dataset; therefore, we had to train the last few layers of the network. This is because the first layer of the inception network extracts the features related to the edges and low-resolution images of the cancer lesion in the MPA cancer dataset. There is no requirement to train the network which already extracted the relevant features. The use of the same weight matrix of inception-resnet-v2 actually helps to initialize the network with proper initialization value to the trainable set of layers [53]. This indeed helped to train the network faster.

Saugata [89] work, done on the same MPA dataset, suggests that A-line infor-

mation is also a useful feature vector for cancer diagnosis. Rajanna et al. [75] also included the frequency content of the PA A-line signal that increased the cancer detection accuracy to 95 % with the same dataset. Therefore, in addition to the wavelength vector, we incorporated the A-line feature vector in our deep learning. We prepared the dataset of 105-channel stacked C-scans consisted of 21-slice C-scan of PA image cube at five wavelengths in include the wavelength and A-line feature vectors to the dataset in precise order. The second part of the thesis was to implement deep 3D CNN. The model was trained on the thyroid because we had more number of thyroid samples. However, the model was able to detect cancer with AUC of 0.85; this is less than previous studies [28, 75]. In previous work, the significant similarity was observed in the image features that defined the cancer region both in thyroid and prostate [28]. We decided to exploit this similarity to handle the limited sample number challenge by implementing the model developed for thyroid but tested on the prostate with the AUC of 0.72, thus supporting our hypothesis and encouraged us to proceed along this line of investigation. Since a deeper network generally improved the predictive performance of the model, we decided to mix the thyroid and prostate datasets for the same cancer detection problem[96].

The third distinguishing factor of this thesis is our attempt to design a deep neural network that is capable of detecting the presence or absence of cancer in a given specimen, regardless of where it is spatially located in the specimen. This is a much more challenging problem than what has been done in studies with the same dataset [75], where ground truth consisted of the definition of cancer lesion

region precisely defined by histopathology slide image of the given specimen. The discriminant analysis in the previous work was performed only on the pixel values in the PA image defined by the pathology slide and all the pixels cancer pixels from all the specimens were pooled together for discriminant analysis. In contrast, the only ground truth that we considered given for our present work was whether a cancer region is present or not present in a given specimen. This challenge was met by implementing the deep 3D CNN with more layers compared to the deep 3D CNN with seven layers [51], doing so requires more samples. The previous result [51] suggested mixing two datasets to increase the number of samples. This is because there is a similar trend when classifying cancer from normal tissue. The dataset was divided into train, validation, and test. The training dataset was still limited but augmented heavily up to 6,200 by rotation, scaling, translation, adding noise, adding/subtracting pixel intensity, random warping, vertical flipping, and horizontal flipping [16]. This is because the number of samples helps to improve the predictive performance of deep learning networks. The network with 11 layers was able to detect cancer with AUC of 0.96 with the comparable performance to the previous study [75].

1.3   Organization of this thesis

- CHAPTER 2: PA Camera Used for Ex-vivo Tissue Imaging: This chapter covers a brief introduction of PA imaging, MPA imaging, the protocol for cancer tissue acquisition, and MPA data acquisition from the thyroid and

prostate ex-vivo tissue specimen.

- CHAPTER 3: PA Features Useful for Cancer Detection: This chapter elaborates the importance of PA imaging for cancer tissue diagnosis with the introduction of the spectral signature and frequency distribution of the recorded PA signals of tissue chromophores.

- CHAPTER 4: Previous work for Ex-vivo Cancer detection: This section describes the previous works related to the cancer tissue detection using the same MPA dataset, and introduces the technique of deep learning for cancer tissue detection.

- CHAPTER 5: Current Work for Ex-vivo Cancer Detection

- CHAPTER 6: Transfer Learning for Cancer Detection: This section implements the inception-resnet-v2 to detect cancer tissue with the AUC of 0.73.

- CHAPTER 7: 3D CNN for Cancer Detection: This method is an extension of Chapter 6 with the use of deep 3D CNN. This model was trained on the thyroid dataset and evaluated on the thyroid and prostate dataset with AUC of 0.85 and AUC of 0.72 respectively. The purpose of this study was to evaluate the ability of PA imaging to extract similar characteristics from two tissue locations such as thyroid and prostate specimens.

- CHAPTER 8: 3D CNN: Cancer Detection and Localization: This method trained on the mixture of the thyroid and prostate MPA dataset with the

deeper network. The network was tested on the mixture of the thyroid and prostate dataset with the AUC of 0.96. This project also implemented a 3D version of grad-cam to localize the cancer tissue region using the test MPA dataset automatically.

- CHAPTER 9: Conclusions and Future Works

- Appendix A: Brain Hemorrhage Classification: The deep 3D CNN has developed for the automatic ICH detection during a real-time CT scan to red flag the urgent cases with a promising result with the AUC of 0.87 at Geisinger Health System (GHS).

The following publications are the outcome of this thesis work.

1 Kamal Jnawali, Mohammad R Arbabshirani, Alvaro Ulloa, Navalgund Rao, and Alpen A Patel. Automatic classification of radiological report for intracranial hemorrhage. In Semantic Computing (ICSC), In 2019 IEEE 13th International Conference on Semantic Computing (ICSC), pages 187-190. IEEE. [49]

2 Kamal Jnawali, Bhargava Chinni, Vikram Dogra, and Rao Navalgund. Transfer Learning For Automatic Cancer Tissue Detection Using Multispectral Photoacoustic Imaging. In Proc. Medical Imaging, 2019: Computer-Aided Diagnosis, volume 10950, page to be published. International Society for Optics and Photonics. [50]

3 Kamal Jnawali, Bhargava Chinni, Vikram Dogra, Saugata Sinha, and Rao Navalgund. Deep 3D Convolutional Neural Network For Automatic Cancer Tissue Detection Using Multispectral Photoacoustic Imaging. In Proc. Medical Imaging, 2019: Ultrasonic Imaging and Tomography. [51]

4 Kamal Jnawali, Mohammad R Arbabshirani, Navalgund Rao, and Alpen A Patel. Deep 3d convolution neural network for CT brain hemorrhage classification. In Medical Imaging 2018: Computer-Aided Diagnosis, volume 10575, page 105751C. International Society for Optics and Photonics, 2018. [48]

5 Kamal Jnawali, John P. Kerekes, and Navalgund Rao. "Comparative study of spectral matched filter, constrained energy minimization, and adaptive coherence estimator for subpixel target detection based on hyperspectral imaging." In Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XXIV, vol. 10644, p. 106441V. International Society for Optics and Photonics, 2018. [52]

6 K Jnawali, B Chinni, V Dogra, and N Rao. Photoacoustic simulation study of chirp excitation response from different size absorbers. In Medical Imaging 2017: Ultrasonic Imaging and Tomography, volume 10139, page 101391L. International Society for Optics and Photonics, 2017. [47]

Chapter 2

PA Camera Used for Ex-vivo Tissue Imaging

2.1　Introduction

In photoacoustic imaging (PAI), the tissue specimen is exposed to a short nanosecond laser pulse with the wavelength in near-infrared (NIR) region [116]. This process results in a local and instantaneous pressure increase within the dominant absorbers [47]. The pressure disturbance propagates as US waves that are then detected by the use of the transducer and used to form a PA image [104]. These US waves or PA waves reflected from the target, are captured by US transducers to produce a sequence of A-line signals as in figure (2.9). The A-line signal is simply the display of the time-dependent response of the transducer generated by the US wave. The

time-gated A-line signals are suitably processed and combined to produce the 2D C-scans PA image as in figure (2.10).
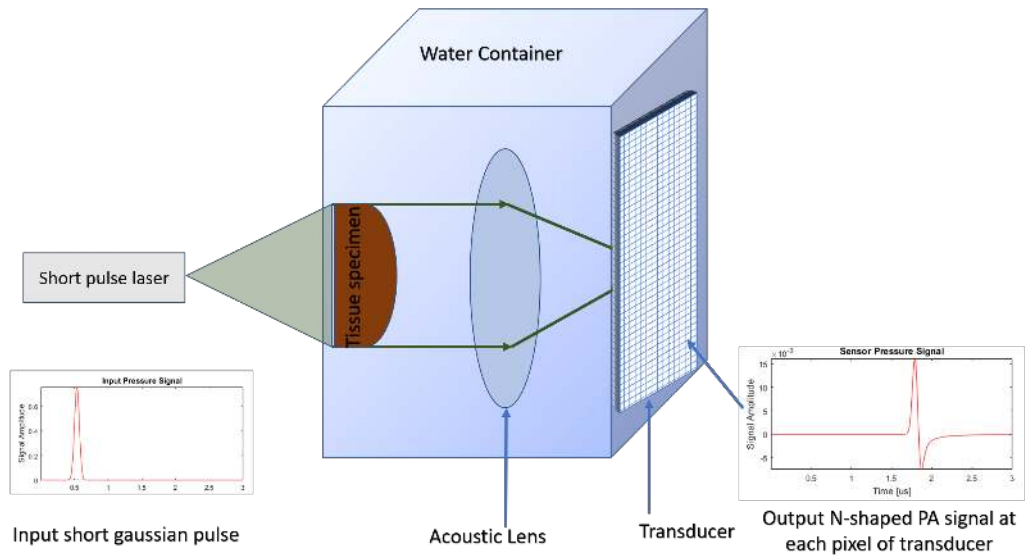
Figure 2.1: Acoustic lens set up with Gaussian short-pulse the input and N-shaped signal as the output. In PAI, the tissue is exposed to a short nanosecond laser pulse with the wavelength in near-infrared (NIR) region. This process results in a local and instantaneous pressure increase within the dominant absorbers. The pressure disturbance propagates as US waves that are then detected by the use of the transducer and used to form a PA image.

During the PA effect, the light wave in tissue converts to ultrasonic waves which scattered much less compared to optical waves [110]. This is because the wavelength of the ultrasonic wave is higher than the light wave. The conversion of the optical to ultrasonic energy helps to bring the following benefits [110].

1 The acoustic waves in the tissue specimen undergo less scattering than the optical wave, according to the Rayleigh scattering law [94]. The acoustic waves

also suffer less attenuation in tissue compared with the light, PAI is capable of providing high-resolution images [85].

2 The photoacoustic imaging (PAI) is capable of imaging the multi-scale tissue specimen with the same contrast [116].

3 The PAI modality is capable of extracting rich details of the optical contrast of the tissue architecture, and molecular distribution as the different optical energy (with the specific wavelength) excites the different tissue chromophores [116].

4 As the amplitude of the PA signal is proportional to the optical absorption of the tissue specimen, the PA image is generally free from background noise and is often speckle-free [116].

5 The ultrasound imaging modality is based on the mechanical contrast of the tissue specimen, and the PAI modality is based on the optical and mechanical contrasts. [103, 116].

2.2 Experimental Set Up

Figure (2.2) shows the actual laboratory set up used for PA imaging of ex-vivo tissues. The tissue specimen was shone with the short pulse at one end, and the generated US waves were focused on the US transducer using an acoustic lens in $4f$ geometry. The tunable laser generates the short Gaussian pulse from the wavelength ranging from 700 to 1000 nm with pulse repetition frequency equals to 10 Hz for PA

imaging [116]. During PA imaging, the sample holder holding the tissue and cylinder containing the acoustic lens and the 32 elements US transducer array were both filled with water [89]. The focal length and diameter of the PA acoustic lens were 39.8 mm and 32 mm, respectively. The depth of field of the lens was approximate ± 0.5 cm around the focal plane. The center frequency of the transducer at another side of the geometrical set up was 5 MHz. The tissue specimen was kept at the $2f$ position from the acoustic lens to avoid the scaling effect. Since the object at $2f$ produces the image at $2f$ with unit magnification [103].
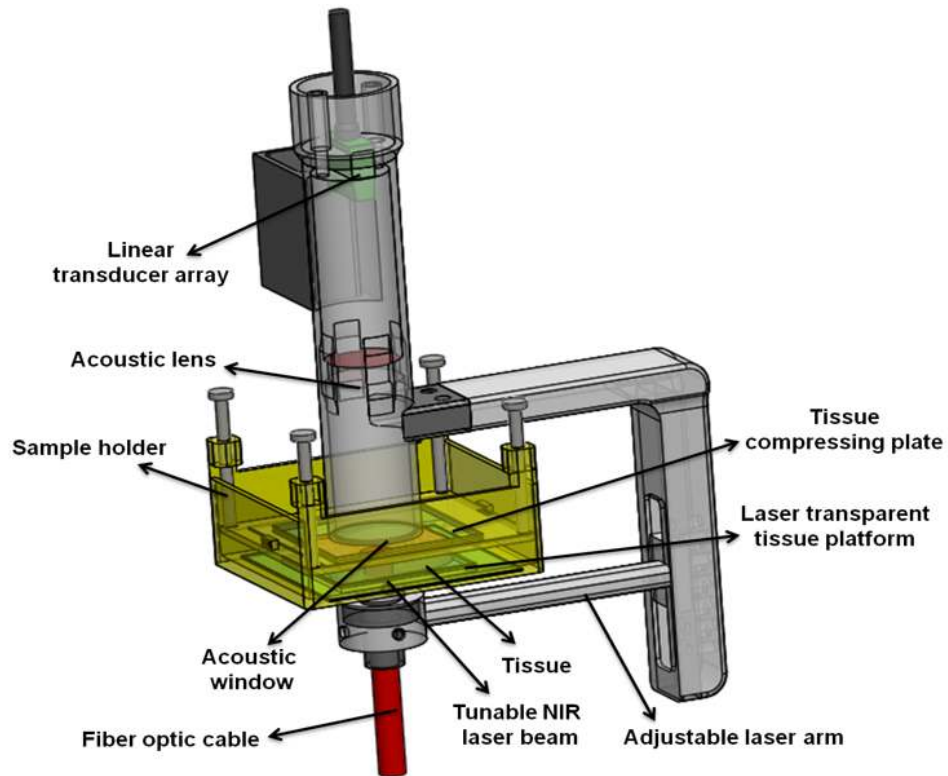
Figure 2.2: Laboratory PA image acquisition set up. Adapted from [90]

The parameters of PA imaging experimental set up is summarized in table (2.1) below.

Table 2.1: PA experiment set up: List of parameters such as frequency, focal length [103]

Description	Value
Center Frequency	5 MHz
Number of elements in transducer	32
Focal length and diameter of acoustic lens	39.8 mm and 32 mm
Bandwidth	60
Pitch	0.7 mm
Depth of field	0.5 cm
Pulse repetition frequency	10 MHz
Wavelength	700 - 1000 nm
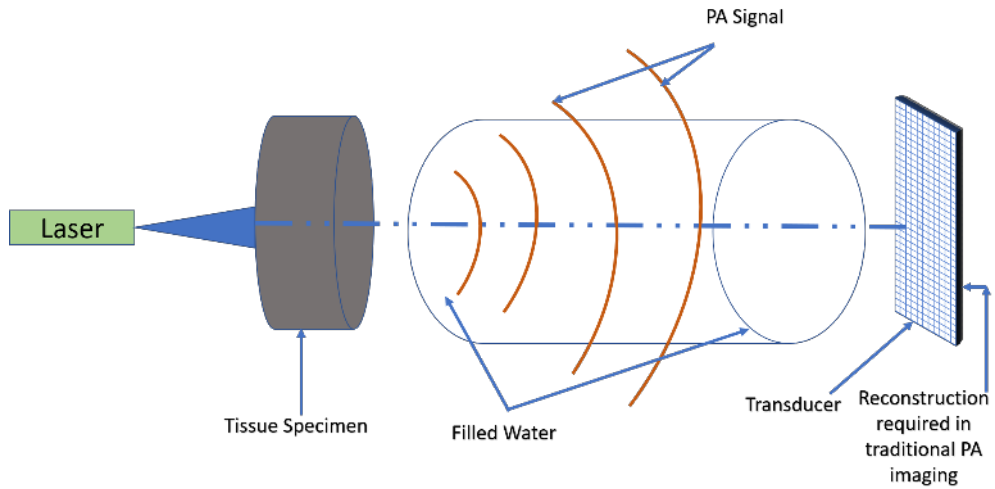
Figure 2.3: Tissue was shone with short-pulse NIR laser light to generate PA waves. Traditional PAI requires the reconstruction to generate PA image. Cox et al. [22] implemented iterative method for the PAI reconstruction.
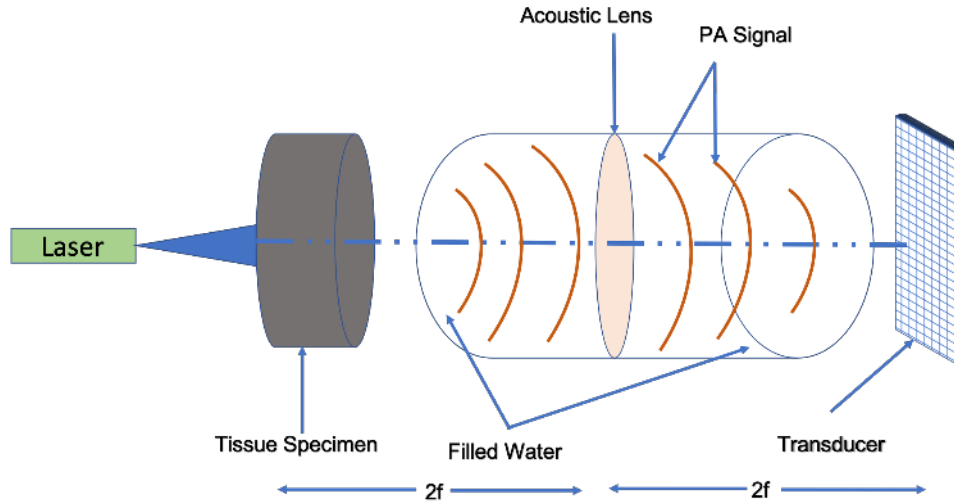
Figure 2.4: Tissue was shone with short-pulse NIR laser light to generate PA waves. The acoustic lens helps to circumvent the reconstruction works. The tissue specimen was placed at $2f$ position and PA image acquired at $2f$ position from the acoustic lens for a unit magnification [32].

2.2.1   Lens based PA imaging acquisition

There is a trade-off between image quality and real-time imaging based on the transducer's number of elements. Furthermore, numerous techniques have been reported for the PA image reconstruction from the recorded PA signals by the transducer. There is a serious mathematical difficulty to use the Radon transform equation for an exact PA image reconstruction [116]. The reconstruction algorithms namely back-projection [41], Fourier based [116], deconvolution based [113] were implemented, however, these algorithms are computationally expensive to generate reconstructed 3D PA image [100]. Rao et al. [76, 77] proposed the acoustic lens, which helps to circumvent the process of the PA image reconstruction [77]. The use of the acoustic lens makes the PA image acquisition process fast, robust, and less error-prone

[77, 103].

PA acoustic lens is similar to an optical lens [77]. A lens helps to converge the scattered wave to a point to produce distinct and high contrast images without computational reconstruction [116]. There is the physics-based correspondence between the optical lens and acoustic lens where the former is used to converge lightwave and later is used to converge acoustic (sound) wave. The biconcave acoustic lens was used to converge the spherically diverging US wave at a transducer when light is traveling from water to lens [20]. The biconvex acoustic lens is used in the optical imaging modality where light travels from air to glass. The object (tissue specimen) was placed at $2f$ location to image the tissue specimen at $2f$ position for a unit magnification [35]. The proper choice of tissue specimen in front of the acoustic lens removes the concern of the scaling effect [20].

2.3 Ex-vivo PA acquisition and ANSI guidelines

PA imaging at five different wavelengths was performed on the freshly excised human prostate and thyroid specimens. The author received the PA dataset prior to the study and wanted to acknowledge [20, 28, 77, 104] for providing datasets. The detailed of the experiment are described below. The thyroid/prostate specimens were collected from confirmed and suspected thyroid/ prostate cancer patients who underwent biopsy. Approval from Institutional Review Board (IRB) was taken along with consent from each patient before conducting the experimental studies [104]. The studies were in compliance with the Health Insurance Portability and

Accountability Act. The excised thyroid/prostate specimen was sent to the surgical pathology laboratory where it was inked and cut into thin sections (2-5 mm thick) by a pathologist after the surgery. One such section with a grossly visible nodule was selected, immersed in normal saline to prevent dryness and sent to PA laboratory to acquire PA images of it at five different wavelengths. Once PA imaging was done, the tissue specimen was returned to the surgical pathology laboratory for histopathology. The entire process to return back the specimen to the surgical pathology laboratory took one hour. It was verified and approved by the pathologist to ensure that the histopathology procedure of the specimen was not compromised [90]. The laser intensity on the excised prostate specimen was maintained around $5 \ mJ/cm^2$, which is well below the safe human exposure limit according to American National Standards Institute (ANSI) guidelines [90]. The workflow of patient protocol is shown in figure (2.5) [28].
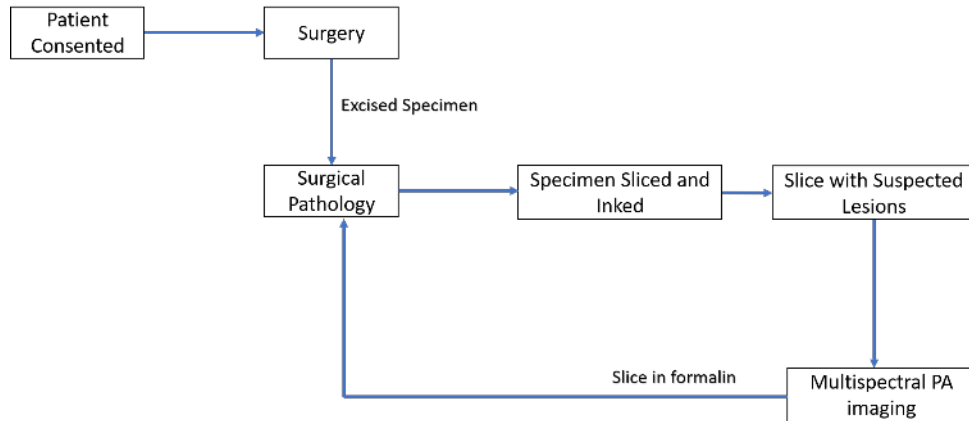
Figure 2.5: Experimental protocol for ex-vivo MPA image acquisition of excised thyroid and prostate tissue samples [28].
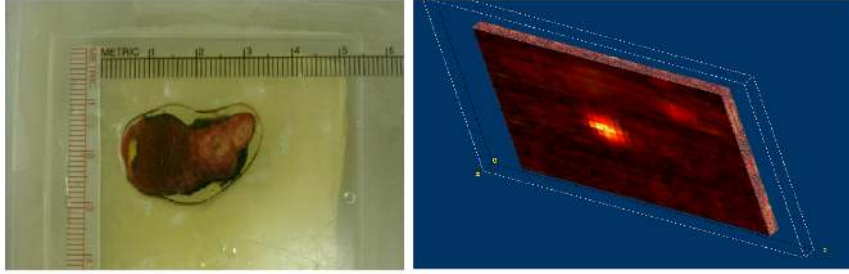
Figure 2.6: Ex-vivo tissue specimen and PA image

It is important to describe the method and the resulting data structure in this section briefly because we used a unique setup to acquire PA signal data set from tissue specimens. The method, the details of which can be found in [20], is illustrated in figure(2.7). An expanded beam of 10 ns pulse of NIR laser light with delivered pulse energy below 20 mJ/cm^2 and pulse repetition rate of 10 Hz was used to expose the tissue specimen immersed in a water-filled medium [47]. PA signals in the form of a short pulse of US generated from absorbers everywhere in the exposed 3D tissue volume propagate towards a specially designed acoustic lens of focal length f. The lens enables the simultaneous focusing of all the waves on the other side of the lens. If the center of the tissue is kept at a distance $2f$ from the lens, then a 32 element linear array of US transducers can be placed at $2f$ distance on the other side to detect the focused PA signals at 32 different pixel locations in the image plane for each laser firing. These US time signals, referred to as A-line signals, were amplified and then digitized at 30 MHz on 32 independent channels simultaneously. A-line signals were envelope detected in order to keep only the slowly varying nonzero signal values. The linear array was scanned in the image plane with repeated laser firing to collect PA signals over the entire image plane. The spatial resolution achieved by this system was around 1.3 mm [90]. PA signals from different depth planes along

the lens axis arrive at the image plane at different arrival times due to the finite

propagation speed of the US in water. By taking time slices on all the A-line signals,

one can generate 2D C-scan PA images that correspond to different depth planes in

the tissue as in figure (2.7). We were able to pick a time gate of 200 sample width,

indicated by t_1 and t_1 in figure (2.7), that included all PA signals coming from the

entire 3 mm thickness of every tissue specimen. The scanning in the image plane

was typically done over a 40×40 pixels with a pixel size of 1 mm by 1 mm. The

typical size of the 3D data was $40 \times 40 \times 200$ pixels.
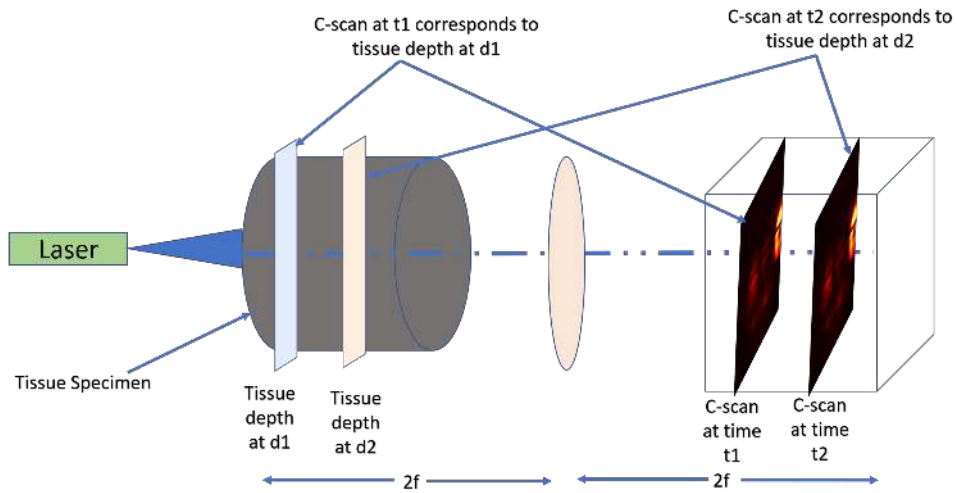
Figure 2.7: Schematic of the PA signal data acquisition setup at one wavelength. An acoustic lens was used to focus the photo-acoustically generated US waves and detect it by a linear US transducer array (not shown in the figure).

2.4   Mathematical Derivation of PA Signal

PA signals originate from optical absorption in three steps. At first a tissue absorbs

light, then absorbed optical energy is converted into heat and produces a tempera-

ture rise, and finally, thermoelastic expansion takes place for the generation of the acoustic wave.

The thermal expansion should be time-invariant to generate acoustic waves [115]. To achieve this requirement, we can either use a short-pulsed laser or a continuous wave laser with intensity modulation at a constant or variable frequency [115]. Upon short laser pulse excitation, local fractional volume expansion $\frac{dV}{V}$ of the heated tissue at position r can be expressed as

$$\frac{dV}{V} = -\kappa p(r) + \beta T(r) \tag{2.1}$$

Here, κ denotes the isothermal compressibility β denotes the thermal coefficient of volume expansion, and p and T denote the changes in pressure (in Pascal) and temperature (in Kelvin), respectively.

For the effective PA signal generation [115].

$$\tau < \tau_{th} = \frac{d_c^2}{4D_T} \tag{2.2}$$

Where τ_{th} is the thermal confinement threshold, d_c is the spatial resolution, and D_T is th thermal diffusivity. Under the assumption of the stress confinement, volume expansion of the absorber during the illumination is neglected, thus we can write this as

$$\tau < \tau_{st} = \frac{d_c}{v_s} \tag{2.3}$$

Where v_s is a speed of sound.

The PA wave generation and propagation in a medium is given [109] by

$$(\nabla^2 - \frac{1}{v_s^2}\frac{\partial^2}{\partial t^2})p(r,t) = -\frac{-\beta}{\kappa v_s^2}\frac{\partial^2 T}{\partial t^2} \tag{2.4}$$

The source term is represented by the right hand side and the wave propagation term is represented by the left hand side in equation (2.4). For the short laser pulse, from equation [2.2],

$$\rho C_v \frac{\partial T(r,t)}{\partial t} = H(r,t) \tag{2.5}$$

Where, H is the heating function and it is related to the optical absorption coefficient μ_a by $H = \mu_a \Phi$ [90].

$$(\nabla^2 - \frac{1}{v_s^2}\frac{\partial^2}{\partial t^2})p(r,t) = -\frac{-\beta}{C_p}\frac{\partial H}{\partial t} \tag{2.6}$$

The generation solution can be shown as [108]

$$(\nabla^2 - \frac{1}{v_s^2}\frac{\partial^2}{\partial t^2})G(r,t;r',t') = \delta(r-r')\delta(t-t') \tag{2.7}$$

The solution of the above equation can be written as

$$G(r,t;r',t') = \frac{\delta[(t-t'-|r-r'|)/v_s]}{2\pi|r-r'|} \tag{2.8}$$

Thus solution to the pressure amplitude is [108].

$$p(r,t) = \frac{\beta}{4\pi\kappa v_s^2}\int dr' \frac{1}{|r-r'|}\frac{\partial^2}{\partial t^2}T(r',t')|_{t'=t-|r-r'|/v_s} \tag{2.9}$$

With additional mathematical step to equation [2.2] [108] results

$$p(r,t) = \frac{r + v_s t}{2r} p_o(r + v_s t) + \frac{r - v_s t}{2r} p_o(-r + v_s t) + \frac{r - v_s t}{2r} p_o(r - v_s t) \qquad (2.10)$$

The first term on the right-hand side represents a converging spherical wave; the second term on the middle side represents a diverging spherical wave that originates from the initially converging wave propagating through the center, and the third term on the left represents a diverging spherical wave. We expect the increase of the pressure amplitude up to maximum value and then the decrease to minimum value [108] as shown in figure (2.8) [47].
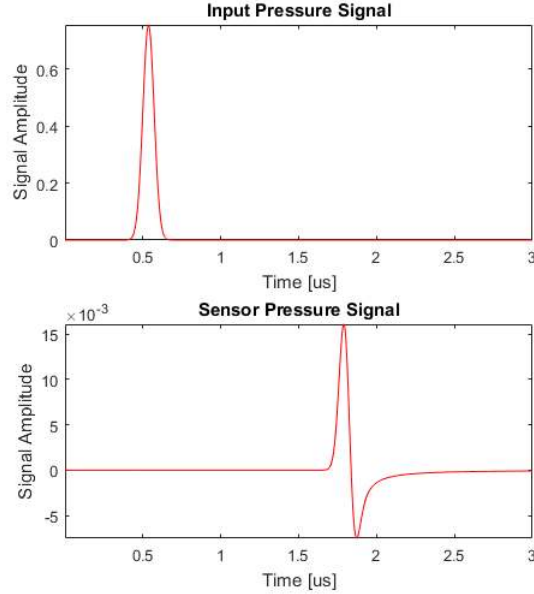
Figure 2.8: The figure shows use of the short pulse of gaussian type as an input and N-shaped PA signal as the output [104].

2.5 Modes of PA imaging

It is useful to discuss the two-mode PA image acquisition, such as the A-line signal and C-scan image in the following section.

2.5.1 A-line PA signal

The A-line is the display of the transducer the voltage-time signal generated by the reflected beam coming from the tissue as in figure (2.9). It is readily observed by connecting the received output to an oscilloscope triggered at the instant of initial excitation. If c is the speed of the US speed in the object is constant, then the time taken by US wave recorded by the transducer is given by

$$\text{time} = \frac{z}{c} \tag{2.11}$$
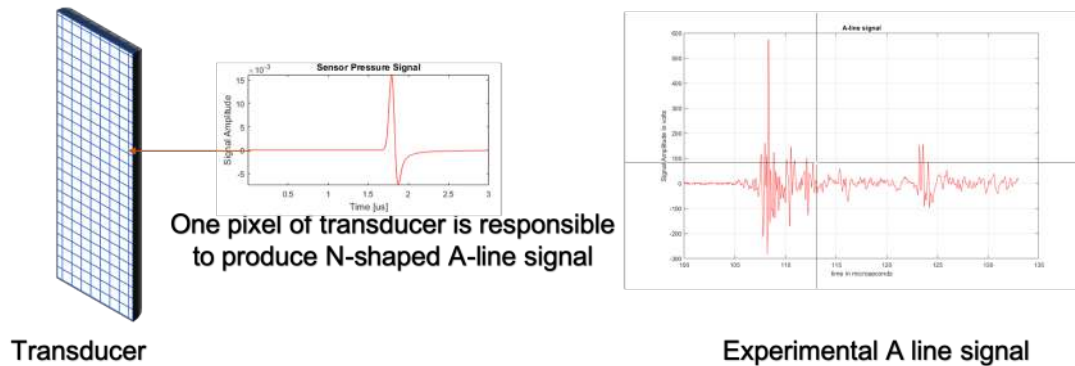
Figure 2.9: The figure shows the theoretical A-mode signal and experimental A-line signal

Where z is the distance.

The A - mode provides information about an object only along the line of sight and within the beamwidth of the transducer [104]. It is tedious and time consuming to move the transducer around the tissue for complete tissue imaging.

2.5.2   C-scan PA image

The C-scan refers to constant depth scanning and responsible for transverse imaging of the tissue, as shown in figure (2.10).

During PA image acquisition, the acoustic lens collected the PA waves generated by PA absorbers in the tissue specimen when shone with short-pulse laser and focused to the point in the image plane [103]. Then the 32 array-sized transducers located at the image plane recorded these PA signals to generate a focused C-scan image with the assumption that wave from all tissue specimen at given depth arrive at the image plane at the same time [103].
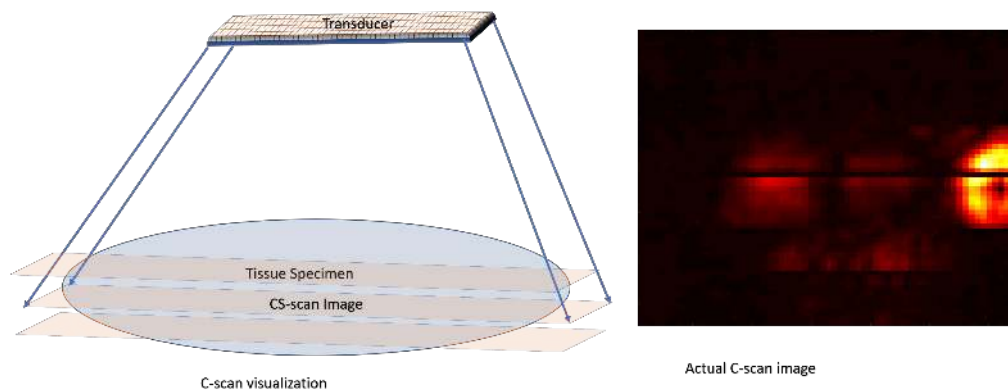
Figure 2.10: The figure shows the theoretical C-scan image and experimental C-scan image

Chapter 3

PA Features Useful for Cancer Detection

3.1 Optical Properties of Tissue Extraction Using PA Imaging

All matters are composed of charged particles such as electrons. The interaction between electromagnetic waves (light) and tissue (charged particles) results mainly two phenomena such as absorption and scattering. In absorption interaction, the absorbed energy can be transferred into heat. The portion of absorbed energy is turned into heat, which is responsible for the photoacoustic (PA) effect [62]. The PA effect is capable of extracting the optical properties of a tissue specimen [104]. The optical properties of biological tissues in the visible and near infra-red (NIR) regions

are related to the molecular constituents of tissue [115]. The optical properties can reveal the changes in angiogenesis and hypermetabolism [116]. The optical absorption spectrum of chromophores (deoxyhemoglobin, oxyhemoglobin, lipid, water) is very sensitive to tissue abnormalities and functions, and hence capable of differentiating normal and cancer tissues. The scattering properties of tissue help to reveal architectural changes in the biological tissue [116]. Both optical absorption and scattering play essential roles in determining the concentration of blood chromophores in tissue [106]. The blood chromophore in malignant tissue absorbs more light than that due to the normal tissue at 760 nm [116], and this may be due to the higher presence of cancer-causing chromophore (deoxyhemoglobin) during the aggressive growth of the tumor [116]. The PAI is a suitable imaging modality to monitor tumor growth and angiogenesis [116]. This is because the underlying physics of the PA imaging technique is based on imaging the optical characteristic of the tissue constituents [109]. That makes the PA imaging modality is capable of functional imaging such as MRI and structural imaging such as US [116].

Table 3.1: Absorption coefficient for chromophores at the five wavelengths

Wavelength	Chromophore	Absorption Coefficient
760 nm	deoxyhemoglobin,	maximum
800 nm	oxyhemoglobin, deoxyhemoglobin	equal
850 nm	oxyhemoglobin	maximum
930 nm	lipid	maximum
970 nm	water	maximum

Figure (3.1) shows the spectral signature of the absorption coefficient of four

chromophores ranging from 650 nm to 1000 nm. The five wavelengths were chosen to extract the maximum information of chromophores [28]. Table (3.1) shows the spectral distribution of deoxyhemoglobin, oxyhemoglobin, lipid, and water with the maximum of the optical spectrum of 760 nm, 850 nm, 930 nm, and 970 nm, respectively. Dogra et al. [28] reported that the PA image at 760 nm is capable of generating a PA image with the brighter pixel intensity distribution corresponding to the oxygen-deficient chromophore (deoxyhemoglobin). This is because the absorption coefficient of deoxyhemoglobin is higher at 760 nm compared to the three other chromophores. The amount of concentration of deoxyhemoglobin is a strong indicator of the presence of cancer [28].
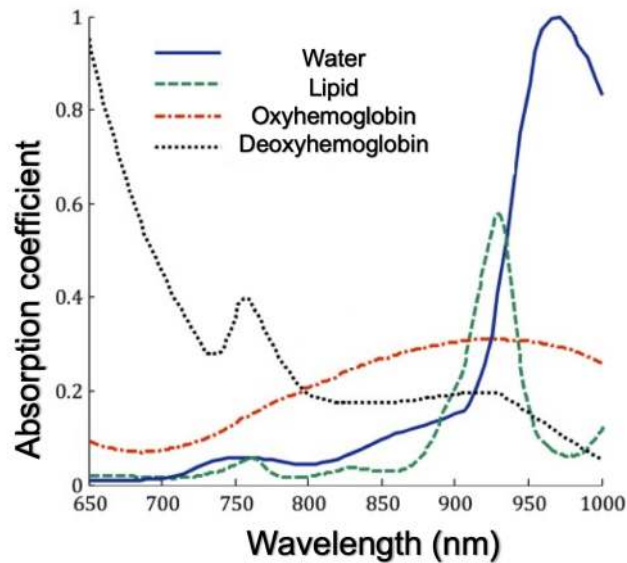
Figure 3.1: Spectral signature of oxyhemoglobin, deoxyhemoglobin, water and lipid. Adapted with permission from [83] © The Optical Society.

3.2   Frequency PA Feature for Cancer Tissue Detection

There is strong evidence of the frequency spectrum dependence on the size of a PA absorber [80, 90]. In tissue imaging, PA absorber represents a tissue specimen that is capable of light absorption. With the assumption that change in the tissue structure and size during a tissue pathology, the frequency content of the PA signal is a significant contributor to differentiate malignant and normal tissue [90]. The spectrum of the emitted PA waves depends on the shape and structures of the chromophores [47].
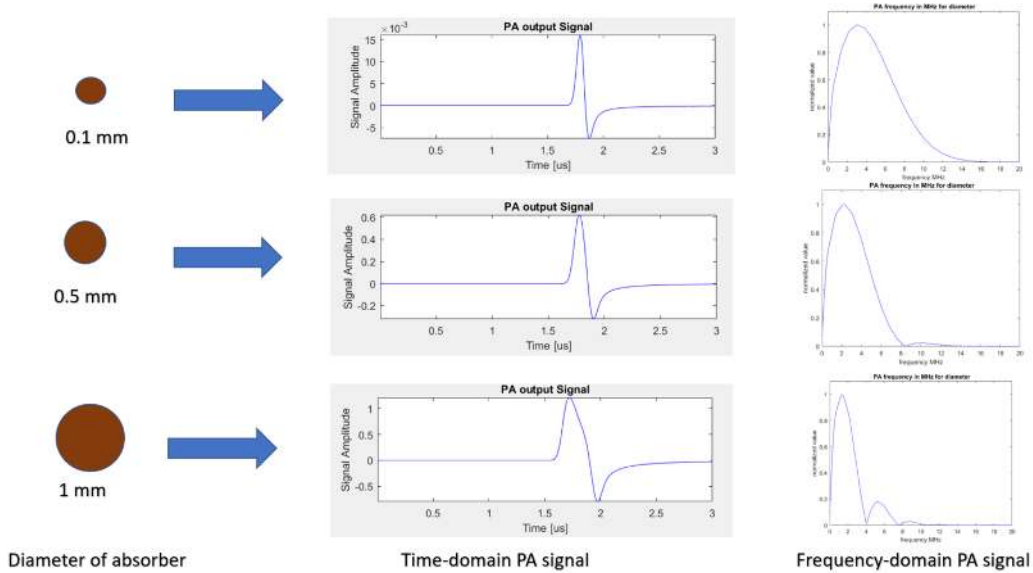
Figure 3.2: Schematic of the dependency of the frequency content of the PA signal with absorber's size with diameter 0.1mm, 0.5 mm, and 1 mm respectively [89].

Xu et al. [116] reported that the malignant tissue is more likely to be in an

irregular shape, and normal tissue is more likely to be in a rounded shape. This may be a contributing factor for the difference in the frequency content of the PA signal in normal and malignant tissue. Sinha et al. [90] reported that frequency content is useful for cancer tissue diagnosis using PA imaging when the size of the tissue specimen is important. Jnawali et al. [47] implemented a k-wave Matlab toolbox [100] to simulate the effect of the size of prostate tissue specimen on the time content (frequency content) of the recorded PA signal.

3.3 Data Preparation

The three-channel dataset with the dimension of the 299×299 pixels were prepared to fit the input of the inception-resnet-v2 network with the use of the first three wavelengths namely 760 nm, 800 nm, and 850 nm. The 105-channel dataset with the dimension of the 64×64 was prepared to fit the input of the deep 3D CNNs proposed in this thesis. The details of the data preparation are described below.

3.3.1 Data preparation: Transfer learning (inception-resnet-v2)

We implemented the transfer learning using inception-resnet-v2 [96] for the cancer tissue detection using the MPA thyroid image dataset. The input MPA data of size $299 \times 299 \times 3$ pixels was prepared by concatenating a one-channel C-scan PA image at three wavelengths, namely 760 nm, 800 nm, 850 nm.
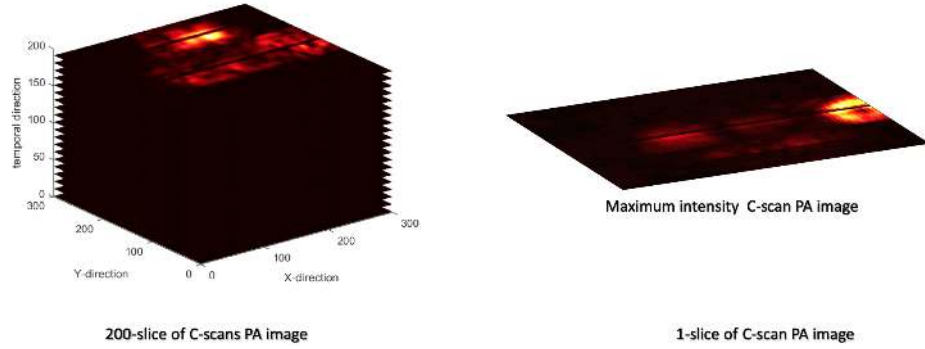
Figure 3.3: The maximum intensity C-scan PA image was extracted from the 200-slice C-scan PA image.
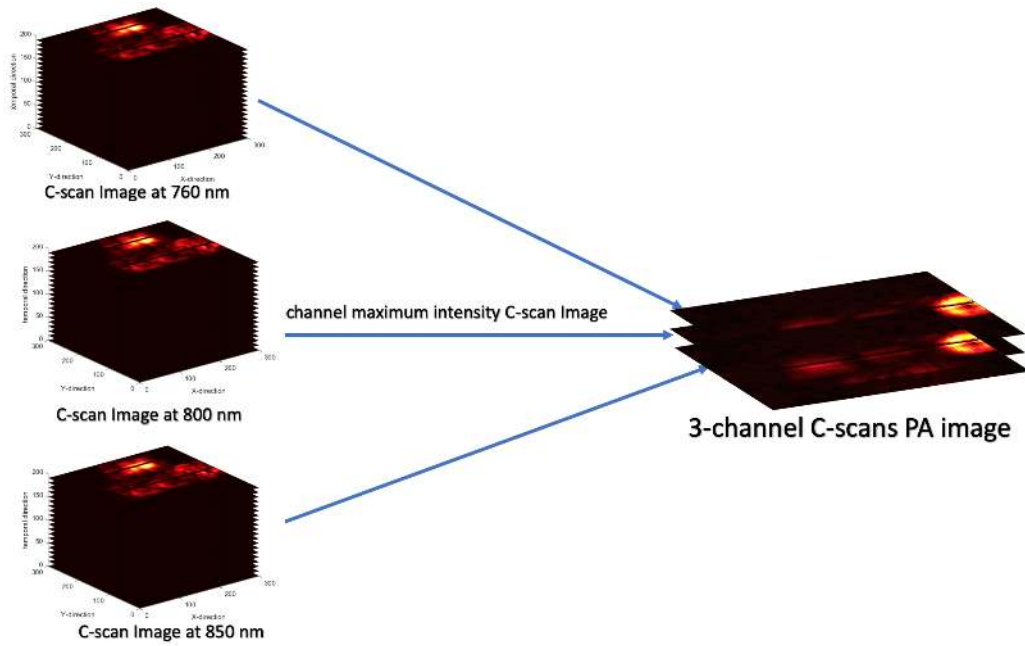
Figure 3.4: Three maximum intensity C-scan of PA images at three wavelengths were concatenated to form 3-channel C-scan MPA image.

3.3.2   Data preparation: Deep 3D CNN

This section describes how the dataset was prepared to incorporate the A-line (frequency) content at the five sets of wavelengths for the deep 3D CNN models. In this

thesis, to incorporate the time domain content at five wavelengths, we stacked the
five-21-slices C-scan 3D PA image to generate a 105-channel 3D MPA image dataset.
The 21-slice of the C-scan PA image was prepared by picking the maximum inten-
sity C-scan from the 200-slice C-scan 3D PA image and then taking its ten adjacent
C-scans. The MPA image at five wavelengths consisting of 105 C-scans is prepared
by concatenating 21-slices of C-scan at five wavelengths as in figure (3.6). The use
of five wavelengths helps to generate a higher contrast PA image of corresponding
chromophores [28]. There is theoretical evidence of the signature of the PA absorber
size that can be seen in the corresponding pulse width of the PA signal. Therefore,
in addition to the wavelength feature vector, the A-line data can be considered as a
possible feature vector [47, 90]. As the size of cancer tissue usually is bigger (more
nodule vessels) than normal tissue and hence could be the stronger PA absorber.
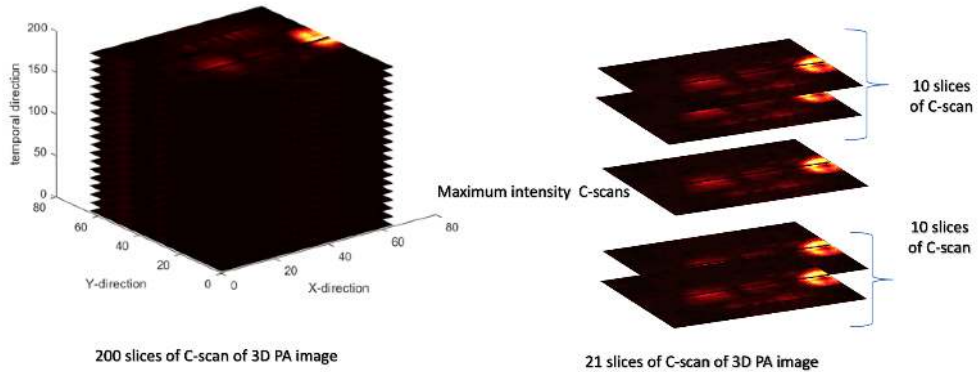
Figure 3.5: Maximum intensity C-scan PA image was extracted from the 200-slice
C-scan PA image, and then took 10 its adjacent C-scans to make 21-slices of C-scan
of 3D PA image.
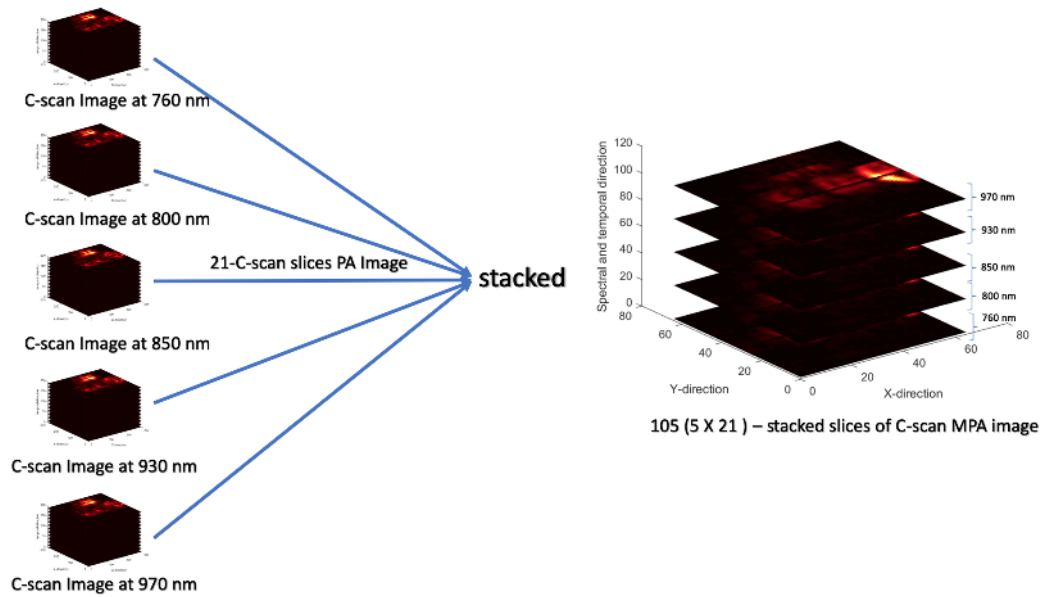
Figure 3.6: The maximum intensity C-scan PA image was extracted from the 200-slice C-scan PA image and then took ten its adjacent C-scans to make 21-slices of C-scan of 3D PA image. MPA dataset was prepared by stacking five 3D PA image cube corresponding to five wavelengths, namely 760 nm, 800 nm, 850 nm, 930 nm, and 970 nm respectively. The 3D PA image cube at each wavelength is responsible for extracting the volumetric distribution of the chromophores (oxyhemoglobin, deoxyhemoglobin, lipid, and fat), and the 3D MPA image cube at five wavelengths is responsible for imaging optically activated chromophores.

Chapter 4

Previous work for Ex-vivo Cancer detection

4.1 Multispectral Photoacoustic Data Acquisition

The PA image cubes of the tissue specimen were taken at five wavelengths (see chapter 3). The size of the PA data cube at each wavelength was ranging from $20 \times 64 \times 200$ pixels to $64 \times 64 \times 200$ pixels. The first two dimensions correspond to the time-gated spatial 2D C-scan image at the given depth, and the third dimension corresponds to the set of time-gated C-scan at the different depth of the tissue specimen. The time information of the recorded PA signal was encoded to the depth using the equation. $d = c \times t$, where d, c, t are depth, the velocity of US, and time respectively.
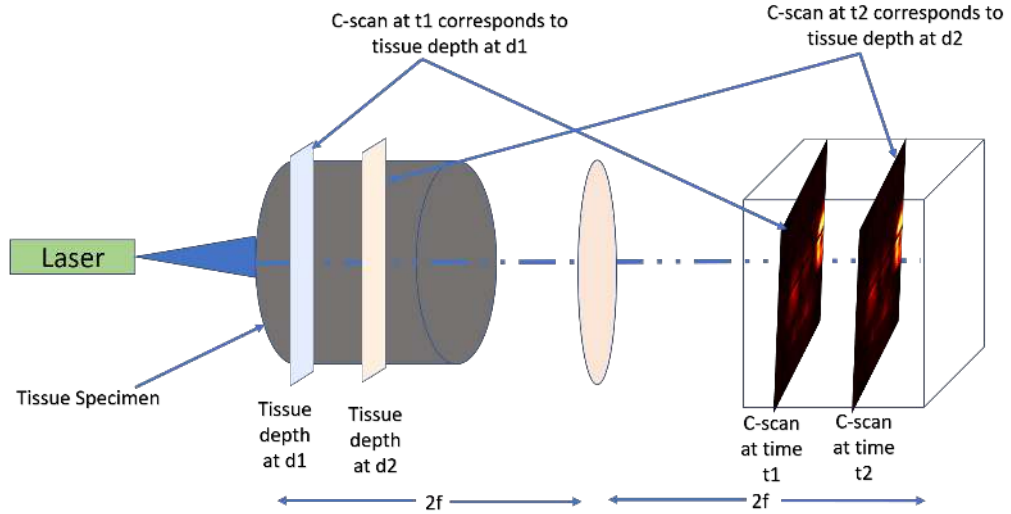
Figure 4.1: Schematic of the PA signal data acquisition setup at one wavelength. The C-scan at t_1, and t_2 corresponds to the tissue specimen at depth d_1 and d_2 respectively. The entire tissue specimen was mapped to C-scans from time t_1 to t_{200}, where t_1 corresponds to depth of depth $4\mu m$ at each wavelength. The five PA image acquisition was made at five wavelengths for each sample. In this paper, the five 21-slice PA image data at five respective wavelengths were concatenated to form a 105-channel MPA dataset.

Each time-gated 2D C-scan corresponds to the tissue specimen of depth $4\mu m$. The multispectral photoacoustic (MPA) dataset was taken at five wavelengths. The set of wavelengths was chosen in such a way to extract the maximum information of the oxyhemoglobin, deoxyhemoglobin, lipid, and water [75, 104]. The PA image corresponding to the tissue specimen also consisted of the histopathological slide and photograph, as shown in figure (4.2). The pathologist provided the marked region corresponding to normal, benign, and cancer in the histopathological slide as in figure (4.2). The samples were divided into cancer, benign and normal groups in previous studies [28, 29, 75, 90]. One sample of prostate and thyroid cancer tissue

specimen photograph, histopathological slide and PA image is shown in figures (4.2) [28, 75], and (4.3) [89] respectively. The malignant region of interest was encircled in the photograph, histopathological slide, and PA image.
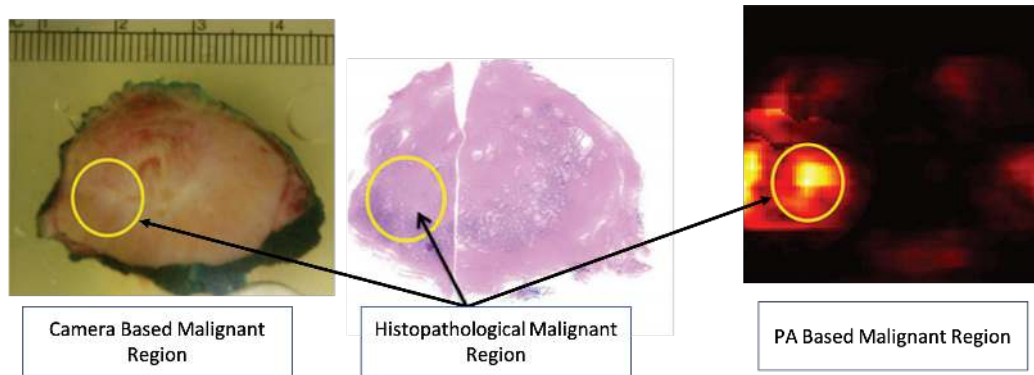
Figure 4.2: Histopathology sample of the prostate specimen with the metric scale, histopathology slide and C-scan PA imaging taken at 760 nm with dimension 45×45. The encircled region with the malignant tissue in the PA image at 760 nm wavelength corresponds to the higher pixel intensity. Reproduced with permission from [28].
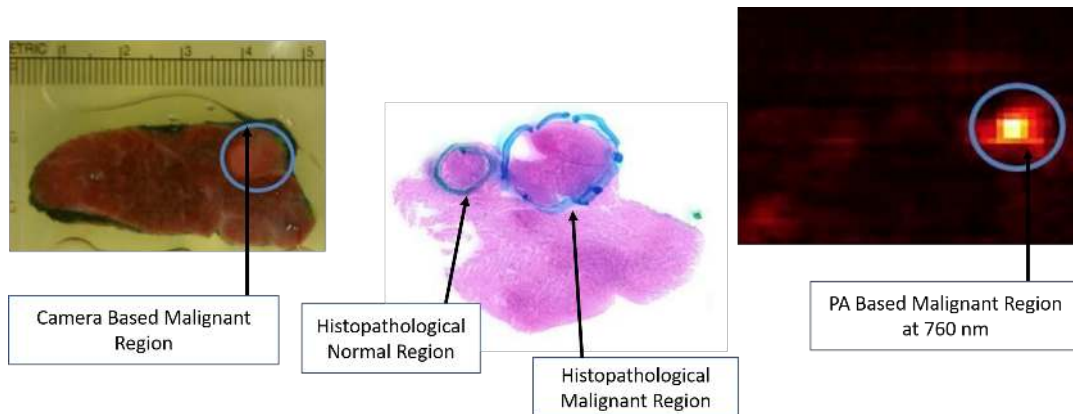
Figure 4.3: Histopathology sample of the thyroid specimen with the metric scale, histopathology slide and C-scan PA imaging taken at 760 nm with dimension 45×45. The encircled brighter region in the PA at 760 nm corresponds to the malignant tissue. Adapted with permission from [90].

4.2 Amplitude and Frequency Based Cancer Tissue Detection Algorithm

Dogra et al. [28], and Rajanna et al. [75] implemented the logistic function and two-layer neural network for cancer tissue detection using manually extracted features. The handcrafted features were extracted by the pathologist and the expert in the field of PA imaging. Dogra et al. [28, 29] implemented the logistic classifiers utilizing the amplitude-based spectral signature using the pixel-based MPA dataset. Dogra et al. reported the accuracy of 0.9 on the prostate MPA pixel-based dataset [28] and accuracy of 0.83 on the thyroid MPA pixel-based dataset [29]. Rajanna et al. [75] implemented the two-layer neural network using the amplitude-based spectral signature and frequency-based PA content of the chromophores with the accuracy of 0.95. The pixel-based MPA samples were extracted from normal, benign, and cancer regions from the PA image (the encircled region in PA image (4.2)). The encircled region corresponding to the normal, benign, and cancer in PA image was generated by the co-registration of the histopathological slide and photograph image. The histopathological slide provided by the pathologist consisted of the normal, benign, and cancerous region. All methods involved the labor-intensive work to co-register the histopathological slide, and photograph to the PA image to extract three regions of interest such as normal, benign, and cancer. Moreover, the process was time-consuming to extract the optimal set of handcrafted features. That demanded the involvement of the expert with domain knowledge such as pathologist and image scientist [28, 29, 75].

Co-registration and Manual Feature extraction

The encircled region required the manual co-registration of the PA image with histopathological slide and photograph(as in figure 4.5), where the pathologist marked the region of interest in the histopathology slide. The process may be an error-prone while extracting the encircled region of interest. The pixel-based samples were used to train, validate, and test the performance of the classifiers. The classifier was not evaluated using the whole MPA image dataset but using the pixel-based samples extracted from the encircled region of interest of the MPA image (figure 4.2).

Figure 4.4: Visualization of the image co-registration involving the photograph-based image, histopathological-based image, and PA-based image.
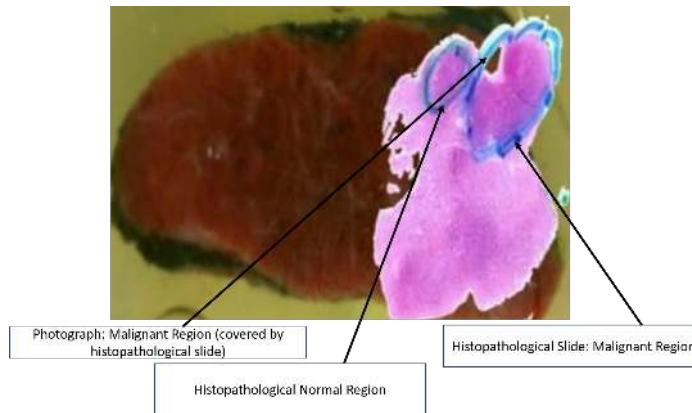
Figure 4.5: The Figure shows the co-registration (overlay of PA image is not shown) method proposed in the previous works. Adapted with permission from [90].

The figure (4.5) shows the selection of the region of interest in the PA image (not

shown) by overlaying the histopathological slide and tissue specimen photograph [89]. The step used for the co-registration is shown in figure (4.4).

A careful choice of handcrafted features improves the predictive performance of a classifier and make more robust, however, it demands the involvement of expert with domain knowledge such as pathologist, and image scientist [28, 29, 75]. The handcrafted features may not always close to perfection; it is possible that a deep neural network itself is capable of extracting features ranging from low level to high level in solving the problem at hand very close to perfection at the hierarchical structure [59].

4.2.1   Amplitude Based Analysis Using MPA Pixel-Based Dataset

Dogra et al. [28, 29] used the handcrafted spectral feature for the cancer tissue detection using the logistic regression model. The authors took a one-slice C-scan image out of 200-slice C-scans of PA image at each wavelength and concatenated at five wavelengths (figure (4.2)). The five-channel C-scan image at five wavelengths was converted to 4 chromophore images (oxyhemoglobin, deoxyhemoglobin, lipid and water) using equation (4.1) [89] as in figure (4.6). Equation (4.1) helps to generate higher pixel intensity corresponding to the malignant tissue in the deoxyhemoglobin channel for the presence of cancer in a given tissue specimen as in figure ((4.7) e and f). Finally, the author extracted the region of interest from chromophore image at five wavelengths by coregistering the PA image with histopathological slide and photograph-based imaging.
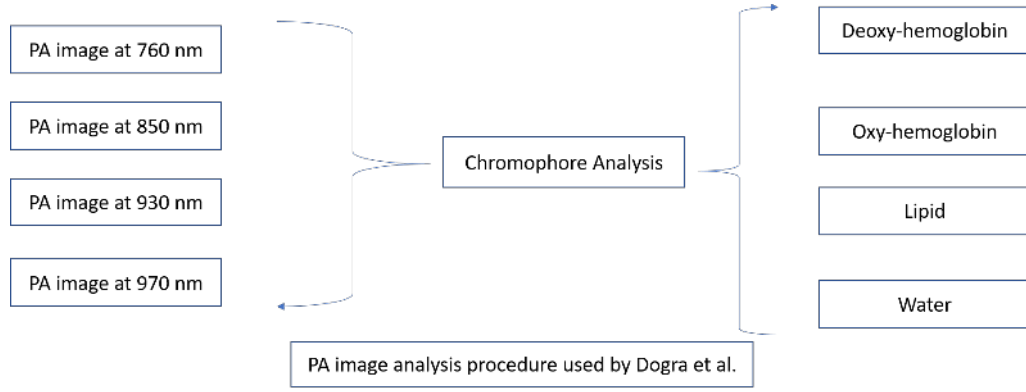
Figure 4.6: Visualization of proposed chromophore analysis [28, 29] with mathematical equation of form described in equation (4.1)

$$
\begin{bmatrix} C_{deoxyhemoglobin} \\ C_{oxyhemoglobin} \\ C_{fat} \\ C_{water} \end{bmatrix} = (M^T M)^{-1} M^T \begin{bmatrix} P^{760} \\ P^{800} \\ P^{850} \\ P^{930} \\ P^{970} \end{bmatrix} K
\tag{4.1}
$$

Where, K is proportionality constant related to the ultrasound parameter [89], and α is absorption coefficient, P^λ is intensity measurement at wavelength λ, $C_{chromophore}$ is the concentration of chromophore and

$$
M = \begin{bmatrix} \alpha^{760}_{deoxyhemoglobin} & \alpha^{760}_{oxyhemoglobin} & \alpha^{760}_{fat} & \alpha^{760}_{water} \\ \alpha^{800}_{deoxyhemoglobin} & \alpha^{800}_{oxyhemoglobin} & \alpha^{800}_{fat} & \alpha^{800}_{water} \\ \alpha^{850}_{deoxyhemoglobin} & \alpha^{850}_{oxyhemoglobin} & \alpha^{850}_{fat} & \alpha^{850}_{water} \\ \alpha^{930}_{deoxyhemoglobin} & \alpha^{930}_{oxyhemoglobin} & \alpha^{930}_{fat} & \alpha^{930}_{water} \\ \alpha^{970}_{deoxyhemoglobin} & \alpha^{970}_{oxyhemoglobin} & \alpha^{970}_{fat} & \alpha^{970}_{water} \end{bmatrix}
\tag{4.2}
$$
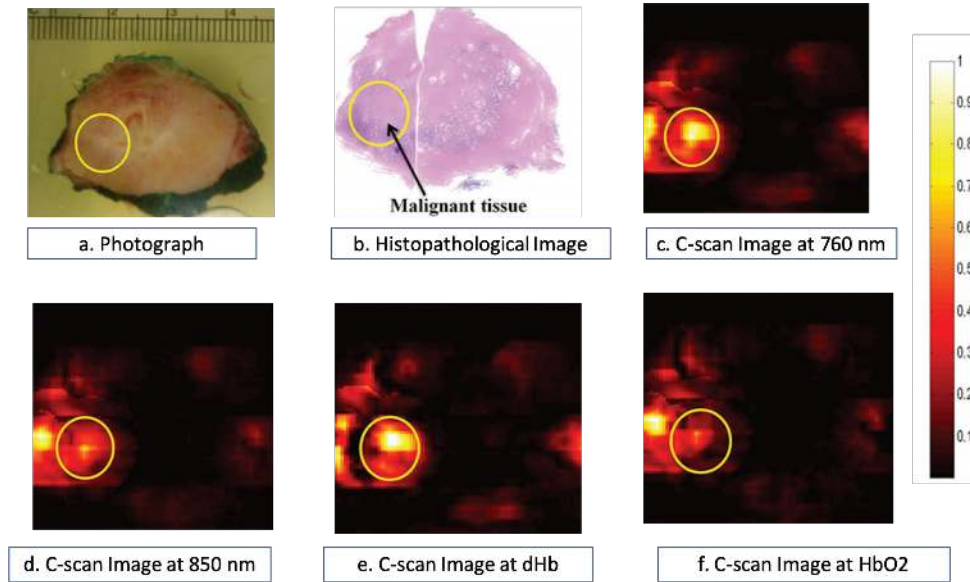
Figure 4.7: a. Tissue specimen, b. Histopathology of prostate with malignant region encircled, c. PA image acquired at 760 nm wavelength, d. PA image acquired at 850 nm wavelength, e. PA image showing absorption of dHb, f. PA image showing absorption of HbO_2. Brighter pixel distribution of dHb is seen in the region of interest corresponding to malignant prostate tissue compared to HbO_2. Reproduced with permission from [28].

The process consists of the following approach to generate the sample dataset. The first step was to convert five wavelength MPA image dataset to four-channel PA image corresponding to oxyhemoglobin, deoxyhemoglobin, lipid, and water using the chromophore transformation equation. The second step was to extract the set of pixels from the encircled region of interest consisting of normal, benign, and cancer. The third step was to take the average of the encircled set of pixels to generate a pixel with four-channels. Furthermore, only the average amplitude-based pixel feature for each sample was used, thus ignoring any possible (2D and 3D) spatially discriminant features as in figure (4.8).

region of interest extracted from the
encircled region

Pixels within the Encircled
region

Average pixel within
the Encircled region
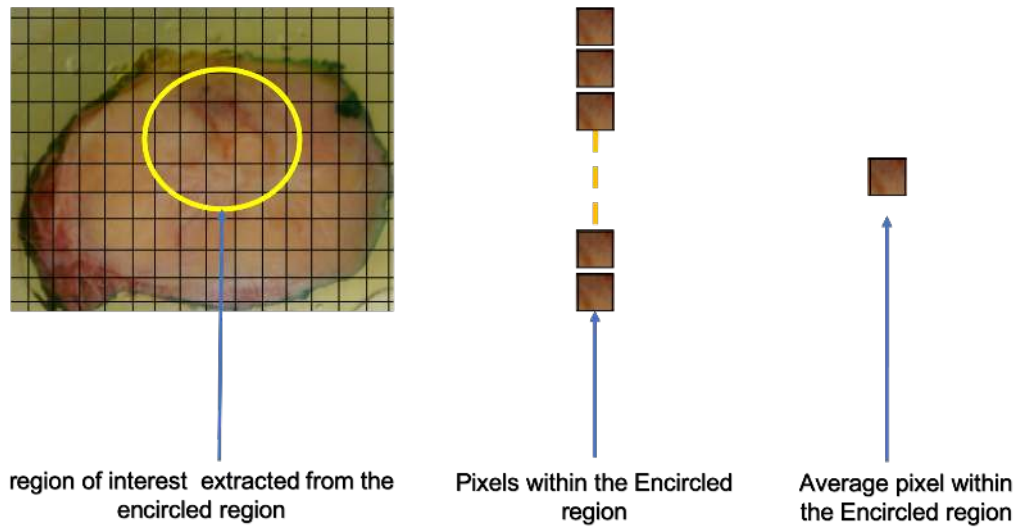
Figure 4.8: Type of samples used in the logistic classifier [28, 29]. The use of pixel-based samples unable to extract spatial information.

4.2.2 Amplitude and Frequency Based Analysis Using MPA pixel dataset

Rajanna et al. [75] implemented the neural network for the cancer tissue detection with the handcrafted frequency and spectral features using the pixel-based MPA samples. Rajanna et al. [75] used 29-dimensional features for each pixel and ranked the most promising features. The 29 features consisted of five PA maximum amplitude of A-line signal at five wavelengths, four chromophore concentrations at four chromophore domain, and five pixel-amplitude intensity distribution of the samples at five wavelengths, and 20 frequency parameters. Rajanna et al. [75] implemented a recursive feature learning algorithm to rank the best feature contributing to the cancer tissue detection. Rajanna et al. [75] reported that amplitude at five wavelengths and frequency contents attribute to the most promising features.

This method also trained and evaluated the performance of the classifier using the set of pixels extracted from the encircled region of interest. Only pixel-based handcrafted features were used; thus, this works also ignores any possible (2D and 3D) spatially discriminant features.

region of interest extracted from the encircled region   Pixels within the Encircled region
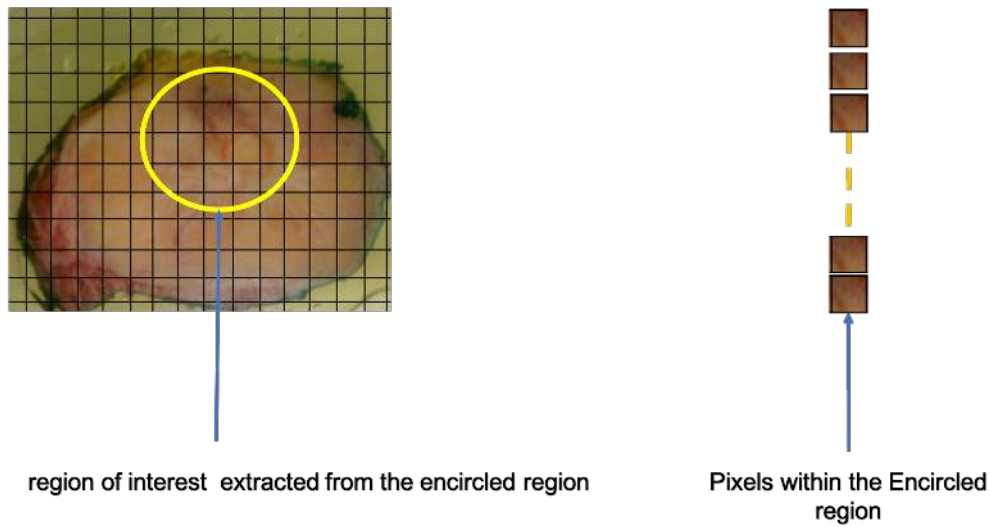
Figure 4.9: This figure shows the type of samples used proposed by Rajanna et al. [75]. The method is not able to extract spatial information for cancer tissue detection.

4.3   Conclusion

Dogra et al. [28] implemented the first cancer tissue classifier using the pixel-based samples extracted from the MPA image dataset. Rajanna et al. [75] improved the classification accuracy with an additional 29 features using the prostate MPA dataset, where five features were common with Dogra et al. studies [28, 29]. The

previous studies [28, 75] trained and evaluated the classifiers using the pixel-based MPA samples extracted from encircled region consisting of normal, benign, and cancer from the whole MPA image dataset with the handcrafted features extracted by the pathologist and expert in the field of PA imaging. The encircled region of interest was extracted by the coregistration of the PA image, histopathological slide, and camera image. The models still require manual work for the coregistration, and not yet become fully automated. Additionally these models also fail to include spatial information while detecting the cancer as in figures (4.8, 4.9).

Chapter 5

Current Work for Ex-vivo Cancer Detection

The machine learning classifiers used in the previous methods were trained, validated, and tested on the MPA pixel-based dataset, not in the whole MPA image dataset. Dogra et al. [28] and Rajanna et al. [75] used pixel-based MPA samples. These processes consisted of the extraction of the pixel-based MPA samples from the encircled region. The encircled part of the PA image was extracted by co-registration of the histopathological slide, photograph image, and PA image. The process was labor-intensive and time-consuming, and yet the classifier is not fully automated.

The current method is trained and evaluated on the whole MPA image, and hence, it removes the labor-intensive work required to extract the pixel-based samples using co-registration during testing. That makes the current method automatic

for detecting the cancer regions in a given specimen[50, 51].

The current method also seeks to extract features automatically using deep learning algorithms from the whole MPA image [48]. The previous studies used pixel-based handcrafted features such as pixel intensity and frequency content[28, 29, 75] but ignored the spatial and volumetric distribution of chromophores (blood). These work fail to extract spatial and volumetric information (figure (8.4)), which becomes essential when cancer tissue extends in all directions [48]. The 2D CNN using the inception-resnet-v2 [96] performs automated extraction of the spatial blood concentration but fails to extract the volumetric information (figure (8.4)). As the cancer tissue spreads in all directions, it is useful to implement the deep 3D CNN [48]. Jnawali et al. [51] improved the performance of cancer tissue detection with the use of the 3D CNN compared to the use of 2D inception-resnet-v2 transfer learning network [50]. The deep 3D CNN is capable of extracting volumetric blood concentration [48, 51] with the 3D convolutional kernel cube, which reduces cost, time of analysis, and improves predictive performance. Though deep 3D CNN was much shallower, the deeper network, in general, is a good choice to build a robust and more predictive classifier [39]. However, the number of samples often limits the implementation of deep neural networks. In the current study, we mix the thyroid and prostate MPA dataset to increase the number of samples to design the deeper 3D CNN for automatic cancer tissue detection and localization with the improved performance compared to the previous studies using deep learning algorithms [50, 51]. The three ways of feature extraction used in previous [28, 29, 75] and current studies

[50, 51] are compared in figure (5.1). The first figure shows the pixel-based distribution, the second figure shows the spatial distribution, and the third figure shows the volumetric distribution of the chromophores.
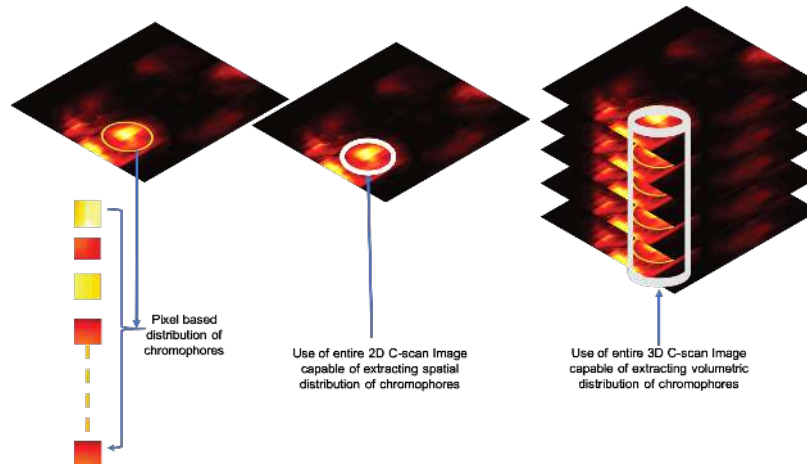
Figure 5.1: The figure shows the type of samples used to train, validate, and, test the performance of the classifiers described in the previous studies [28, 29, 75] and current studies [50, 51]. The left figure shows the pixel-based samples used in the previous studies [28, 29, 75]. This shows the requirement of the co-registration of PA image with the histopathological slide and the photograph of tissue specimen to extract the encircled region of interest during testing for a cancer tissue detection. Thus the method used in the previous studies still requires manual work for co-registration during test time. This thesis proposed three models to detect cancer in the given specimen using an entire C-scan image, and hence no need of manual work as required in the previous works even during training. The middle figure shows the entire C-scan PA image, which was used in the transfer learning network. This method was able to extract the spatial distribution of the chromophores as shown by the circle while using the entire C-scan image. The right figure shows the concatenated 21-channel entire C-scan PA image which was used in the two 3D CNNs [51]; deep 3D CNN with seven layers and deep 3D CNN with 11 layers. This data structure was able to extract the volumetric distribution of the chromophores with more predictive performance than that using only the spatial distribution. These two later processes become automated with the use of an entire MPA image during the test time for the cancer tissue detection. That is different from previous works [28, 29, 75].

5.1 Current works

The current work is divided in to three parts, the first part describes the implementation of the transfer learning on the thyroid cancer tissue detection, the second part describes the implementation of the deep 3D CNN on thyroid and prostate cancer tissue detection, and the third part describes the implementation of the deeper 3D CNN on the mixture of thyroid and prostate cancer tissue detection with localization.

5.1.1 Automatic Feature Extraction, Detection, and Localization of Cancer Tissue Using Deep Learning

We implemented 2D transfer learning using inception-resnet-v2, deep 3D CNN with seven layers, and deep 3D with 11 layers for the cancer tissue detection using the thyroid, thyroid, and mixture of thyroid and prostate MPA image dataset respectively. The same dataset was used in previous studies [28, 29, 75], which develop cancer tissue classifiers using handcrafted features.

Transfer learning

We implemented the transfer learning using inception-resnet-v2 [96] for the cancer tissue detection using the MPA thyroid image dataset. The input MPA data of size $299 \times 299 \times 3$ pixels was prepared by concatenating a one-slice C-scan PA image at three wavelengths, namely 760 nm, 800 nm, 850 nm.

Deep 3D CNN

We implemented the two architecture of deep 3D CNNs for cancer tissue detection using the MPA thyroid image dataset, and a mixture of an MPA thyroid and prostate image dataset [51]. The input MPA data of size $64 \times 64 \times 105$ pixels was prepared by concatenating a 21-slice C-scan PA image at five wavelengths, namely 760 nm, 800 nm, 850 nm, 930 nm, 970 nm.

Automatic Localization

We implemented the grad-cam algorithm [84] using the deep 3D CNN for the automatic localization of cancer region in a test MPA image.

5.2   Conclusion

While previous efforts using the same dataset involved decision making using mathematically extracted image features, this work demonstrates that this process can be automated without any significant loss of accuracy. Another major contribution of this work has been to demonstrate that both prostate and thyroid datasets can be combined to produce improved results for cancer diagnosis with the automatic localization of the cancerous region.

Chapter 6

Transfer Learning for Cancer Detection

This work is taken from the paper published by SPIE Medical Imaging, 2019. *[Kamal Jnawali, Bhargava Chinni, Vikram Dogra, and Rao Navalgund. Transfer Learning For Automatic Cancer Tissue Detection Using Multispectral Photoacoustic Imaging. In Proc. Medical Imaging 2019: Computer-Aided Diagnosis, volume 10950, page to be published. International Society for Optics and Photonics]* [50].

Abstract

Pathology diagnosis is usually done by a human pathologist observing tissue stained glass slide under a microscope. In the case of multi-specimen study to

locate cancer region, such as in thyroidectomy, significant labor-intensive processing is required at high cost. Multispectral photoacoustic (MPA) specimen imaging, has proven successful in differentiating photoacoustic (PA) signal characteristics between a histopathology defined cancer region and normal tissue. A more pragmatic research question to ask is, can MPA imaging data predict, whether a sectioned tissue slice has cancer region(s)? We propose to use inception-resnet-v2 convolutional neural networks (CNNs) on the thyroid MPA data to evaluate this potential by transfer learning. The proposed algorithm first extracts features from the thyroid MPA image data using CNN and then detects cancer using the softmax function, the last layer of the network. The model achieved an area under curve (AUC) of cancer, benign nodule and normal are 0.73, 0.81, and 0.88 respectively.

6.1 Introduction

Currently, deep learning tasks are being used in the field of digital pathology, and are successful in cancer (lymphoma) classification tasks [46]. The number of the samples in the dataset to solve a medical imaging problem are often limited due to many constraints. However, deep learning architectures demand significant number of samples in a given dataset for a robust predictive performance. To overcome this problem related to limited samples, Litjens et al. [65] discussed the possibility of transfer learning technique in the medical imaging field. Transfer learning is a technique where a deep learning architecture learns the features from one domain, for example, natural image and later apply the technique to another domain which can be different but deals with the realted problem. Transfer learning also helps

to train the network faster than training the network from a scratch. It also helps to minimize over-fitting issues [55]. Khosravi et al. [58] successfully implemented the transfer learning to cancer classification tasks. In recent years [48], CNNs were used as a popular classifier for an automatic feature extractor and classifier. In this paper, we implemented a transfer learning technique to the limited MPA image thyroid dataset using the state of the art inception-resnet-v2 [96] deep learning algorithm. The performance of the model was evaluated by calculating the area under the curve (AUC) of the receiver operating characteristic (ROC) curve [36], precision recall, and f_1 score [6].

6.2   Dataset preparation and analysis

The PA images of each thyroid samples were generated with the short-pulsed laser at five different wavelengths: 760 nm, 800nm, 850 nm, 940nm, and 970 nm [28]. We prepared the MPA dataset using the first three wavelengths 760 nm, 800 nm and 850 nm. The spatial dimensions of the PA image in the dataset were ranging from 20×64 pixels to 64×64 pixels. In the depth direction; we have considered 200 slices correspond to 5 mm of tissue thickness. The single-spatial-2D C-scan image with maximum pixel intensity was extracted from the 200-spatial-2D C-scan at each wavelength. Finally, one MPA dataset was prepared by concatenating three single-spatial-2D C-scan image at three different wavelengths 760 nm, 800nm, 850 nm respectively. The spatial dimension of the concatenated image was increased to 299×299 pixels using bilinear interpolation to fit into the input of the inception-

resnet-v2 network. One sample of thyroid cancer tissue specimen and its PA is show in figure [6.1].

Thyroid specimen   **Histopathology slide**   **PA image cube**
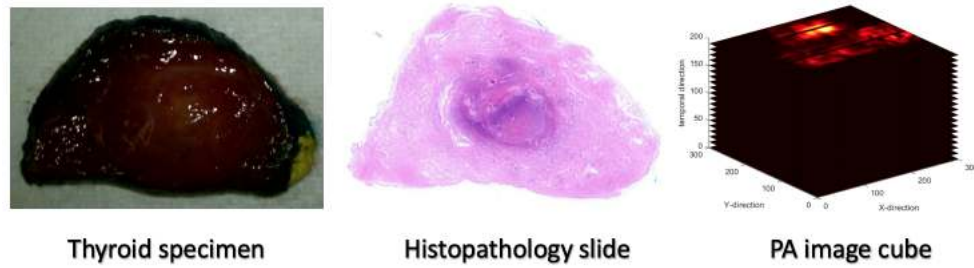
Figure 6.1: Sample of the thyroid specimen with the metric scale, histopathological slide and 3D PA image of the specimen of size $299 \times 299 \times 200$, where one pixel of spatial 2D image is equal to 0.7×0.7 mm^2.

The data structure used to train the inception-resnet-v2 is shown in figure [6.2].
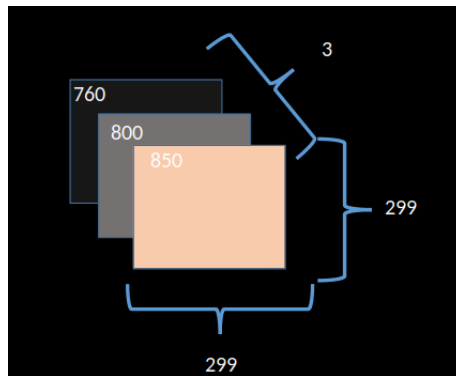
Figure 6.2: The structure of the input MPA dataset with size $299 \times 299 \times 3$ pixels fed to the inception-resnet-v2.

Table 6.1: Normal, benign and malignant tissue distribution of thyroid MPA dataset

Tissue Specimen	Thyroid
Normal	40
Benign	46
Cancer	17
Total	103

6.3   Experiment

In this paper, 103 samples, as shown in table (6.1) of thyroid specimens were taken to train, validate, and test the inception-resnet-v2 network for the cancer tissue detection. The MPA dataset were annotated into three groups: normal (40 specimens), benign nodule (46 specimens) and cancer tissue (17 specimens). The dataset was divided into train (73 specimens) and test dataset (30 specimens) in the ratio of 0.7 : 0.3 respectively, and the train dataset was further divided into the ratio of 0.8 : 0.2 into train (58 specimens) and validation dataset (15 specimens) respectively. The model was trained using five-fold cross validation [15] where the dataset is divided into five groups and one group corresponds to validation set and the rest groups correspond to training set. The cross validation method while training a classifier is useful to minimize the variance in the model [15].

The inception-resnet-v2 [96] network for the cancer tissue detection was trained in tf-slim, a lighter version of tensorflow in python. The inception-resnet-v2 network is a variation of inception-v3 network [97] borrows ideas from the residual network [38]. The data was first divided in to the train, validation and test data set, and

then converted to a tfrecord data structure.

In this paper, we changed the number of nodes in the final layer of inception-resnet-v2 to three for three class classification. The network was initialized with the image net weights [96] and trained with learning rate initialized at 0.0001 and decayed by the factor of 0.1 in each two epochs. The network weights were updated with adam optimizer [57] during back propagation [42]. The MPA dataset was augmented by rotation, translation, mirroring, vertical and horizontal flipping, and cropping. The image augmentation technique helps to increase the number of sample to design an efficient deep neural network [59]. The network was trained for the 100 epochs while saving the best classification score to the disc using the model checkpoint [2] module. The module provides a method for the early stopping method along with an option to monitor metric during training. The early stopping [74] with the patience of 5 was chosen while training the network. The dropout [93] of 0.9 was applied to the network to prevent overfitting. These automated methods help to train the network with improved performance and reduced overfitting.

6.4 Result

The performance of the model was evaluated by generating the classification reports, and AUCs of ROC curve for the cancer, benign nodule and normal tissue. The classification report with precision, recall, and f_1 score is shown in table [6.2]. The AUCs of the ROC curve for the three classes are shown in figure [6.3]

	Precision	Recall	f_1 Score	Support
Cancer	0.80	0.50	0.62	8
Benign Nodule	0.69	0.75	0.72	12
Normal	0.75	0.90	0.82	10
Avg	0.74	0.73	0.72	30

Table 6.2: Precision, recall, f_1 score and support
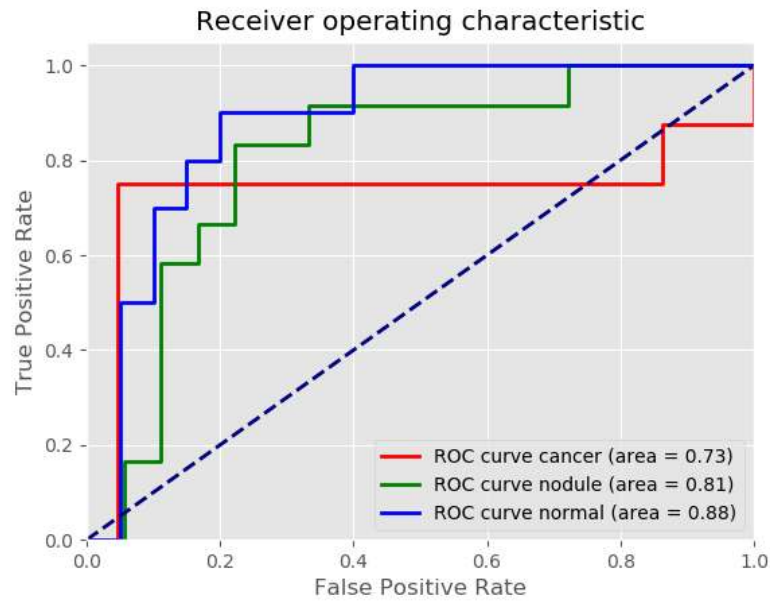
Figure 6.3: ROC curve for the cancer, benign nodule and normal thyroid tissue specimen respectively compared against each other

6.5   Conclusion and Future works

In this article, we implemented the deep neural network based on transfer learning with the inception-resnet-v2 for the cancer tissue detection with limited MPA

dataset with very promising results (AUCs for cancer, benign nodule and normal are respectively 0.73, 0.81, and 0.88). To the best of our knowledge, this is one of the first examples of the use of a deep transfer learning to cancer tissue detection using PA imaging. The use of the inception-resnet-v2 network as the deep transfer learning tool for the limited dataset shows a potential for cancer tissue detection even when the dataset was often limited. We found that the prediction performance of the network is more robust to benign nodule and normal tissue than cancer tissue. This may be due to the greater number of normal and benign tissue compared to cancer tissue.

The primary future direction of this work is to collect a significant number of balanced MPA datasets to improve the predictive performance of the network, as a predictive performance of a deep learning architecture improves generally with the greater number of samples [48]. Jnawali et al. [47] reported that PA signal is strongly dependent on the shape and size of the tissue. The secondary future direction of this work is to develop the deep 3D CNN [51] to the existing dataset to extract the volumetric temporal (frequency) tissue information. Jnawali et al. [48] reported that a 3D CNN will be better choice than a 2D CNN when volumetric information becomes important.

Chapter 7

3D CNN for Cancer Detection

We implemented deep 3D CNN to detect cancer tissue using the MPA thyroid image dataset with improved performance compared to the transfer learning network described in chapter 6. This chapter also seeks to evaluate the ability of MPA imaging to extract the characteristic features corresponding to cancer tissue in the prostate tissue specimen; irrespective of a tissue specimen location.

This work is taken from the paper published in SPIE Medical Imaging, 2019. *[Kamal Jnawali, Bhargava Chinni, Vikram Dogra, Saugata Sinha, and Rao Navalgund. Deep 3D Convolutional Neural Network For Automatic Cancer Tissue Detection Using Multispectral Photoacoustic Imaging . In Proc. Medical Imaging 2019: Ultrasonic Imaging and Tomography.]* [51].

Abstract

Multispectral photoacoustic (MPA) specimen imaging modality is proven successful in differentiating photoacoustic (PA) signal characteristics from cancer and normal region. The oxy and de-oxy hemoglobin content in human tissue captured in the MPA data are the key features for cancer detection. In this study, we propose to use deep 3D convolutional neural network trained on the thyroid MPA dataset and tested on the prostate MPA dataset to evaluate this potential. The proposed algorithm first extracts the spatial, spectral, and temporal features from the thyroid MPA image data using 3D convolutional layers and detects cancer tissue using the logistic function, the last layer of the network. The model achieved an area under curve (AUC) of 0.85 and 0.72 on the thyroid and prostate MPA dataset respectively.

7.1 Introduction

The American cancer society estimates about 164,690 new cases of prostate cancer- the third leading cause of cancer- which is estimated to cause about 29,430 deaths in the United States for the year 2018 [1].

Photoacoustic (PA) imaging [116] is an emerging, noninvasive, functional and molecular imaging modality that exploits the PA effect to combine the strength of contrast of optical imaging and spatial resolution and penetration depth of the ultrasound imaging [28]. In PA imaging, tissue is irradiated by a short-pulsed laser light in the near-infrared region. When light is locally absorbed, heat is generated. The generated heat converts into increased pressure. The increased pressure propagates

as an ultrasound wave and generates the PA signal which is finally recorded by a transducer [47]. The PA signal amplitude is proportional to the optical absorption coefficient of tissue [28]. The PA image contrast is based on the distribution of optical absorption and scattering of all tissue chromophores (deoxyhemoglobin, oxyhemoglobin , lipid, water) [28]. Multispectral photoacoustic (MPA) imaging is capable of containing oxy and deoxy hemoglobin at multiple wavelengths, a key feature for cancer tissue detection. This has been demonstrated in previous studies [28, 29].

Recently convolutional neural networks (CNNs) are becoming very popular due to their capacity to tackle complex computer vision learning problems. CNNs are used in number of computer vision problems such as image classification tasks [48, 59], caption generation [105], text classification [49] and image segmentation [24]. In past few years, deep learning algorithms have been successfully used for the digital pathology [46] for classification tasks. The power of CNN lies in its capability to extract the discriminant features at multiple level of abstraction without human involvement. CNNs are very appealing because of their ability to extract a hierarchy of increasingly complex features specific to a problem at hand.

In this article, the deep 3D CNN was designed for the cancer tissue detection using the thyroid and prostate MPA dataset. The same dataset was used in previous studies [28, 29, 75] for cancer tissue detection based on the spectral, and frequency handcrafted features. In the previous studies, the authors [28, 29, 75] trained and tested their classifiers only on the region of interest (cancer/benign/normal), ex-

tracted from the MPA dataset. The region of interest in the PA image was extracted by co-registration of PA image and histopathology slide as in figure (7.1) where the region of interest in the histopathology slide was marked by the pathologist. Dogra et al. [28, 29] used the handcrafted spectral feature for the cancer tissue detection using a logistic regression model. Rajanna et al. [75] implemented a neural network for the cancer tissue detection with the handcrafted frequency and spectral features using the 807 samples of pixel taken from the 42 prostate MPA dataset. Rajanna et al. [75] used 29 features and ranked the most promising features. The previous methods ignored the spatial features. The estimation of the concentration of chromophores at multiple wavelength helps to measure the blood volume [11] for cancer tissue detection. Since the malignant tissue is generally richer in blood volume than normal tissue [72], including blood volume measurement and spatial features in cancer tissue detection improves a predictive performance of the network. Previous studies used manual extraction of spectral [28, 29, 75] and frequency (temporal) [75] features but ignored blood volume and spatial features. This CNN performs automated extraction of blood volume and spatial features in addition to automated extraction of spectral and temporal features with the 3D convolutional kernel cube, reduces cost, time of analysis and improves predictive performance.

In this paper, we introduced the deep 3D CNN as an automatic spatial, spectral and temporal(frequency) cancer tissue feature extractor and detector. The current method reduced the very time consuming manual work of feature extraction and co-registration of the histopathology slide to PA image. In addition to the detec-

tion, this paper also evaluated the ability of the MPA imaging technique to extract common prominent features of cancer in the two tissue location as the network was trained with the thyroid MPA dataset and evaluated on the thyroid and prostate MPA dataset.

7.2 Dataset

In this paper, the samples were divided into non-cancer and cancer while in the previous studies the samples were divided into cancer, benign and normal tissue group [28, 29, 75, 90]. The ground truth was annotated by a radiologist. The size of the data cube at each wavelength was ranging from $20 \times 64 \times 200$ pixels to $64 \times 64 \times 200$ pixels. The first two dimensions capture the spatial 2D C-scan image and the third dimension captures the volumetric temporal A-line along the depth direction where one C-scan slice corresponded to tissue of depth $4\mu m$. One sample of prostate cancer tissue specimen and its PA image is shown in figure [7.1].
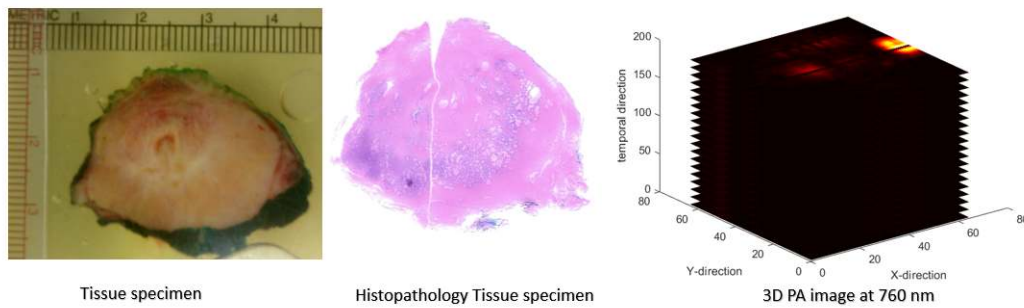
Figure 7.1: Histopathology sample of the prostate specimen with the metric scale and 3D PA image of the specimen of size $64 \times 64 \times 990$
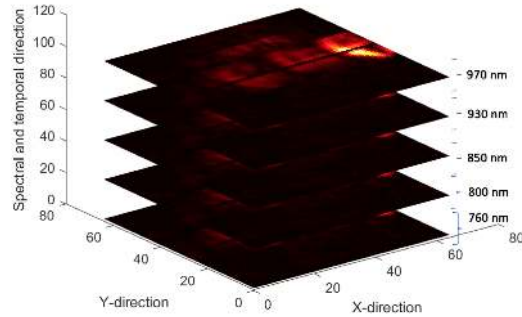
Figure 7.2: The structure of the input MPA dataset with size $64 \times 64 \times 105$ pixels fed to the proposed deep 3D CNN architecture.

The 21 spatial 2D C-scans were extracted from the 200 spatial 2D C-scan PA image from each wavelength of each sample. Then spatial dimension of each PA sample was increased to 64×64 pixels with a bilinear interpolation to make the size of all dataset uniform. Finally, each sample data was prepared by concatenating five 21-2D PA image cube corresponding five wavelengths 760 nm, 800 nm, 850 nm, 930 nm and 970 nm respectively to 105 C-scan MPA datacube as in figure (7.2).

7.3   Method

In this paper, the proposed architecture consisted of two 3D convolutional layers, two 3D max pooling layers, two fully connected layers and a classifier as shown in figure (8.7). The network was trained up to 25 epochs with batch size of 80 while saving the model corresponding to the lowest classification error. The network was initialized with the He initialization parameter [39] that makes back-propagation efficient. The batch normalization [44] was applied after each convolutional layer, and fully connected layer for a faster training. The batch normalization also acts as a regularizer to improve the network's predictive performance. Dropout of 0.5

was applied after each fully connected layer that acts as regularizer[93] to prevent a over-fitting. Rectified linear unit (ReLu) activation function was applied after each convolutional layer and fully connected layer for efficient back propagation by reducing the gradient vanishing problem [59]. The network's model parameters were updated with the Adam optimizer [57] during the back propagation. The network was trained with a learning rate equal to 0.01 and decayed exponentially in each epoch by 0.1. The hyperparameter set {number of convolutional layers, size of convolutional filters, stride size, maxpooling layers, number of fully connected layers, number of nodes in the fully connected layer, activation function, dropout, and learning rate} were optimized with grid search. The first convolutional layer (colored in yellow) consisted of 16 filters with kernel size $3 \times 3 \times 3$ with 1 pixel stride in two spatial X and Y direction and 2 pixels stride in depth direction. The maxpooling layer (colored in green) was used to reduce the number of features from the first convolutional layer. The pooling size of the first maxpooling layer was $3 \times 3 \times 3$. The second convolutional layer consisted of 24 filters with size $3 \times 3 \times 3$ with 2 pixels stride in all directions. The second convolutional layer was followed by the maxpooling layer and the pooling size of the second maxpooling layer was $2 \times 2 \times 2$. The two fully connected layers with 500 and 100 nodes respectively added to the network. Finally, a logistic function was applied to the network for binary cancer tissue detection.
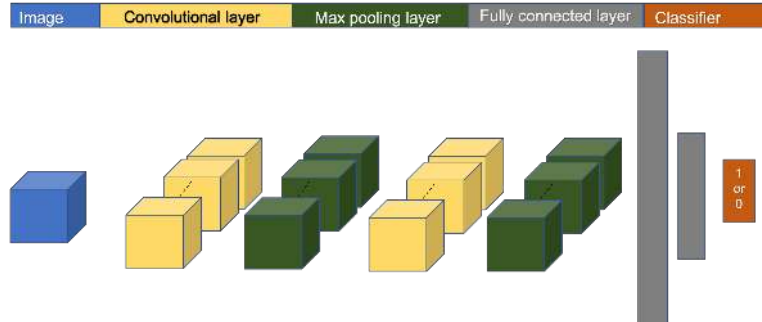
Figure 7.3: The detail architecture of the proposed network for the cancer tissue classification using the MPA dataset

	Input size	Output size
Image size	$64 \times 64 \times 105$	$64 \times 64 \times 105$
Convolutional layer 1	$64 \times 64 \times 105$	$16 \times 64 \times 64 \times 53$
Maxpooling 1	$16 \times 64 \times 64 \times 53$	$16 \times 21 \times 21 \times 17$
Convolutional layer 2	$16 \times 21 \times 21 \times 17$	$24 \times 11 \times 11 \times 9$
Maxpooling 2	$24 \times 11 \times 11 \times 9$	$24 \times 5 \times 5 \times 4$
Fully connected layer 1	$24 \times 5 \times 5 \times 4$	500
Fully connected layer 2	500	100
Binary cancer tissue classifier	100	1

Table 7.1: Hyperparameters used in the deep 3D CNN

7.4 Experiment

The 108 PA thyroid MPA dataset was used to train the 3D deep neural network and the network was tested on the thyroid and 28 prostate MPA dataset. The train thyroid dataset was divided into the train (92 specimens) and test data (16 specimens) in the ratio of 0.85:0.15 respectively. The train data was further divided into the 0.8:0.2 for the training (74 specimens), and validation (18 specimens) dataset respectively. The network was trained with five fold cross-validation [15]. The cross-

validation method while training a classifier is useful to minimize the variance in the model [33]. The network was trained, and validated using the thyroid MPA dataset. The number of samples were increased using image augmentation [59] before training the network. This increases the samples and is useful when dataset is limited. In this chapter, the MPA dataset was augmented by random pixel shift from (1, 10), random pixel rotation from (1, 10), random noise from (1, 10), vertical flipping and horizontal flipping as in figure (7.4). The cancer MPA dataset was augmented almost five times more than the normal MPA dataset. The distribution of data was remained imbalanced after augmentation. In addition to augmentation, a class-weight parameter [118] was introduced while training the network. The validation dataset was used to optimize the hyperparameters of the network. The optimal choice of hyperparameter improves the predictive performance of a deep learning architecture. The set constituted the set of hyperparameters. A grid search method was used to optimize the hyperparemeter for the proposed model. The set of hyperparameters used for the model is tabulated in table (7.1). Finally, the network for the cancer tissue detection was tested on the thyroid and 28 prostate MPA dataset by generating the area under curve (AUC) of receiver operating characteristic (ROC) curve.
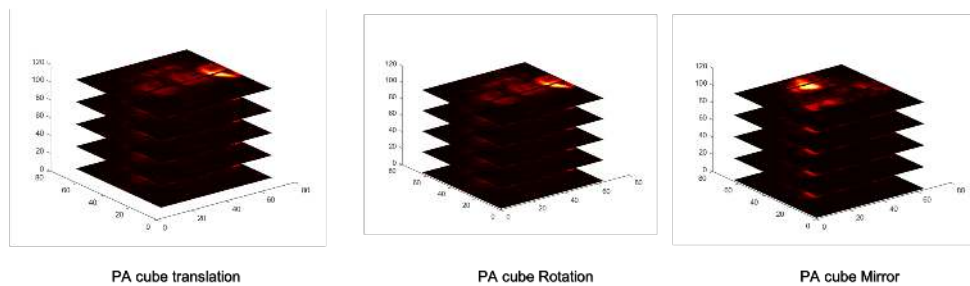
Figure 7.4: Methods used for the data augmentation

7.5   Result

The model was tested on the thyroid and prostate MPA dataset. The AUC with value 0.85 and 0.72 for the thyroid and prostate dataset respectively are shown in figures (7.5) and (7.6).
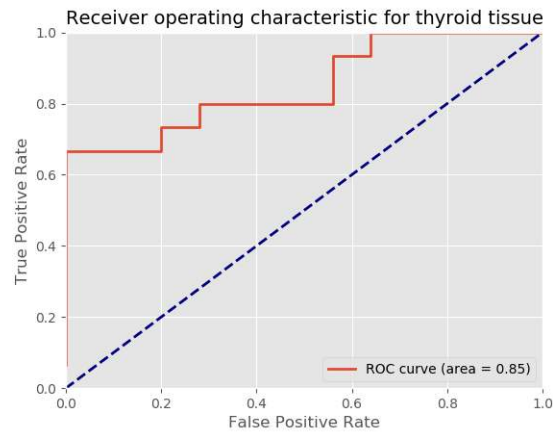
Figure 7.5: ROC curve generated by the model on the thyroid test dataset with AUC of 0.85
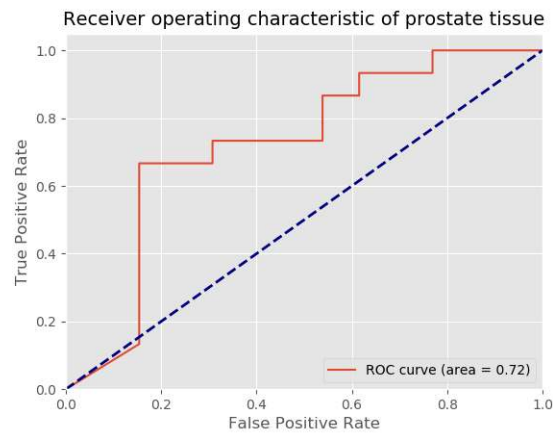
Figure 7.6: ROC curve generated by the model on the prostate test dataset with AUC of 0.72

7.6 Conclusion and Future works

In this article, we implemented a 3D CNN, trained with thyroid MPA dataset to detect thyroid cancer with 0.85 AUC and prostate cancer with 0.72 AUC. We trained the network on the thyroid dataset because of grater number and tested on the prostate. The number of training samples was limited and the distribution of dataset was highly skewed toward normal tissue. To address this problem, data augmentation and class weight[118] methods were implemented while training the network. Furthermore, ROC was chosen as the metric to test the performance of the classifier.

Training the model on the thyroid MPA dataset, and testing on the prostate MPA dataset with promising result may suggest the oxy and deoxy hemoglobin, and frequency content of the PA signal are the prominent features for cancer tissue detection which has been proven [28, 29, 75]. In addition to spectral and frequency content, the model was able to extract the volumetric signature of blood [11] with the cubic convolutional kernels from the MPA dataset. Given the AUC of 0.85 on the thyroid MPA dataset, the AUC of value 0.72, generated by the network, on the prostate MPA dataset, may suggest the similar distribution of the optical, temporal (frequency), and volumetric signature of blood in the thyroid and prostate cancer tissue. These result may indicate the model's ability to extract common prominent features, irrespective of sample location, with a high degree of accuracy. To our knowledge, this work is one of the first applications of a deep 3D convolutional neural network trained on the thyroid MPA dataset and tested on the prostate MPA dataset.

Recently, histopathology slides were used for the prostate cancer diagnosis with very promising result [66] with more number of the sample compared to this study, however the two models were implemented on two different domain dataset. In general the performance of a deep learning architecture improves with the number of samples [48]. The primary future direction of this work is to collect a greater number of tissue samples. Jnawali et al. [50] developed the thyroid cancer tissue detector using the 2D inception-resnet-v2 [96] with a promising result. Since the volumetric time domain [47] and spectral information [29] are promising features to detect the cancer tissue, the secondary future direction of this work is to modify the existing inception-resnet-v2 [96] deep neural network to 3D neural network.

Chapter 8

3D CNN: Cancer Detection and Localization

This model is an improvement over the previous model described in chapters six and seven. This model was trained and evaluated on the mixture of prostate and thyroid dataset with the deeper network. This chapter also implemented a grad-cam algorithm to localize the cancer tissue in the given test MPA dataset automatically.

Abstract

Pathology diagnosis is usually done by a human pathologist observing stained tissue glass slide under a microscope. In the case of multi-specimen study to locate cancer regions, such as in thyroidectomy, and prostatectomy, significant labor-intensive processing is required at a high cost. Multispectral photoacoustic (MPA) specimen

imaging, has proven successful in differentiating photoacoustic (PA) signal charac-
teristics between a histopathology defined cancer region and normal tissue. This
is mainly due to its ability to efficiently map oxyhemoglobin and deoxyhemoglobin
content from MPA images, key features for cancer detection. A fully automated
deep learning algorithm is purposed, which learns to detect the presence of can-
cer in freshly excised ex-vivo human thyroid and prostate tissue specimens using
three dimensional (3D) multispectral photoacoustic (MPA) dataset. The proposed
automated deep learning model consisted of the CNN architecture, which extracts
spatially co-located features, and a logistic function that detects a cancer tissue.
The area under the curve (AUC) was used as a metric to evaluate the predictive
performance of the classifier. The proposed model detects and locates cancer tissue
with the AUC of 0.96.

8.1   Introduction

Siegel et al. [88] reported 595,690 deaths out of 1,685,210 new cases of cancer in
the United States in the year of 2016. 1,980 deaths out of 64,300 reported new
cases of thyroid cancer, and 26,120 deaths out of 180,890 reported new cases of
prostate cancer in the same year [88]. American Cancer Society recommends men
older than 50 years for (prostate) cancer screening [28]. Screening is done with
prostate-specific antigen and digital rectal examination, but both screenings suffer
from low specificity (the percentage of actual negatives that are correctly classified)
and sensitivity (percentage of actual positives that are correctly classified) [28, 88].

   Early detection and risk assessment play a significant role in the success of treat-

ments to improve the survival of patients [28]. Ultrasound (US) imaging is widely used in cancer tissue [87] screening because of its relative safety, low cost, noninvasive nature, real-time display [67]. Trans-rectal ultrasound (TRUS) is a common US imaging for prostate cancer detection; however, there are types of cancers not visible to TRUS [28]. This is because the US has a relatively low contrast, which is based on the detection of mechanical properties of the tissue [45]. Incorrect diagnosis of the cancer tissues using US imaging could lead to delayed diagnosis and treatment or over-diagnosis and over-treatment [101] because of low-resolution structural imaging [28]. In addition to the US, magnetic resonance imaging (MRI) is another imaging for cancer tissue screening; however, it has not been widely used because of its high cost and slow imaging speed [89]. X-ray computed tomography (CT) offers high tissue penetration depth and excellent spatial resolution; however, it suffers from low sensitivity for lesion detection [25]. Gonzalez et al. [25] also reported that X-ray radiation possesses risk of radiation associated with cancers [45]. Sinha [89] reported that CT has minimal application in cancer tissue screening because of its ionizing radiation properties. Positron emission tomography (PET) is capable of extracting functional information such as blood metabolism for cancer tissue detection [70]; however it suffers from low spatial resolution [45]. Single-photon emission computed tomography (SPECT) shares similar physics of imaging technique with PET, but it also suffers from lower spatial resolution [89]. Currently, the primary medical imaging modality used for cancer diagnosis are US, CT, MRI, PET, and SPECT. Although most of these technologies are well established and widely used

in practice, there are problems related to low sensitivity and specificity and higher cost for cancer diagnosis [43].
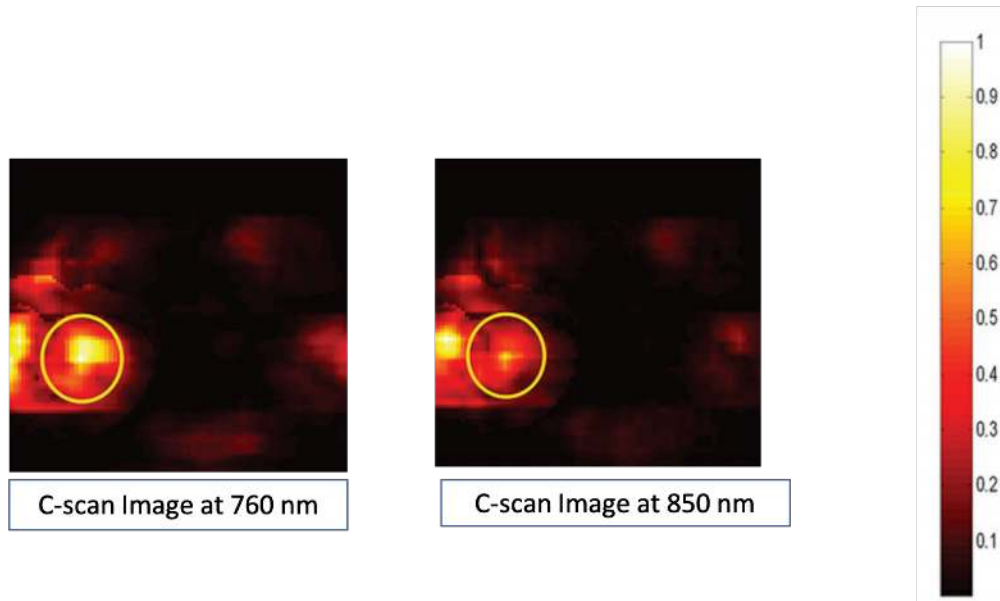
Figure 8.1: The figure shows the pixel intensity distribution of the encircled region of interest of the PA images taken at two wavelengths. The figure shows that the PA image is capable of detecting cancer because of high concentrations of the deoxyhemoglobin at 760 nm compared to the lower concentrations of oxyhemoglobin at 850 nm. Reproduced with permission from [28].

Photoacoustic imaging (PAI) is a new medical imaging technique [104] that is currently making a transition from bench to bedside, both in terms of technology [104] and clinical applications [28]. PAI is based on the photoacoustic(PA) effect, which is a phenomenon of generating acoustic waves from an object illuminated by a pulsed laser light [104]. PAI is an emerging noninvasive soft tissue medical imaging modality that exploits the PA effect to combine the strength of optical imaging is

capable of producing high contrast imaging and ultrasound imaging which is capable of producing high resolution in deep tissue imaging [47, 103]. PA imaging is safe due to its non-ionizing radiation properties [104]. Furthermore, it is also low-cost, which is a huge plus over other imaging techniques like MRI or CT [34]. MRI or CT is not capable of detecting cancer tissue in the early stage and is capable of detecting cancer tissue when the diameter grows to 1 cm in size [107].

The PAI modality is capable of extracting rich details of the optical contrast of the tissue architecture, and molecular distribution as the different optical energy (with the specific wavelength) excites the different tissue chromophores [116]. The PA imaging at the given wavelength is capable of generating high contrast PA images of optically active corresponding chromophore as the chromophore has different absorption properties in different laser wavelengths (higher intensity pixel distributions of deoxy in 760 nm and lower for that of oxy in 850 nm as in figure (8.1). Multiple high contrasts five wavelengths can generate PA images of the optically active chromophores such as oxyhemoglobin (high in oxygen content), deoxyhemoglobin (low in oxygen content), lipid, and water: 760 nm, 800 nm, 850 nm, 930 nm, and 970 nm [28]. These chromophores make the PA imaging system to image optical biomarkers for cancer tissue detection [28, 90]. In multispectral photoacoustic (MPA) imaging, the large difference in light absorption coefficient between blood and other tissue constituents enables detection of tissue angiogenesis associated with rapid tumor growth in early stages [61]. That makes the PA imaging capable of functional imaging. The PA image acquisition produces robust and less error-prone co-registered

images capable of both structural and functional imaging compared to current imaging techniques such as MRI with US [3]. The functional, and structural information such as spatial distribution, volumetric distribution, and the spectral signature of the chromophores can be extracted with the use of multispectral 3D PA imaging techniques [90, 116].
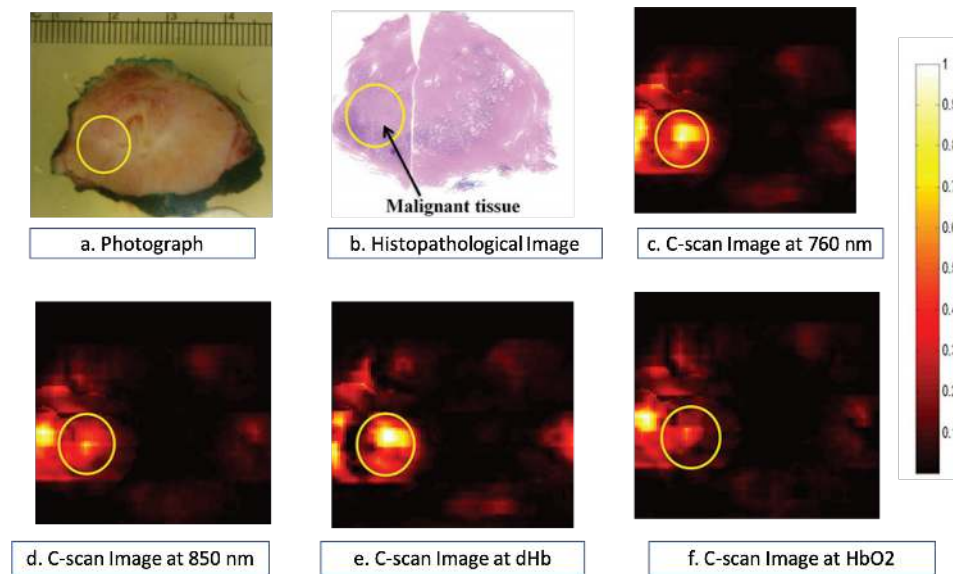
Figure 8.2: a. Tissue specimen, b. Histopathology of the prostate with malignant region encircled, c. PA image acquired at 760 nm wavelength, d. PA image acquired at 850 nm wavelength, e. PA image showing absorption of dHb, f. PA image showing absorption of HbO2 (Reproduced with permission from [28]). The encircled region taken at the deoxyhemoglobin channel of the PA image (figure e) shows the presence of malignant region because deoxyhemoglobin absorbs more light to generate a higher pixel intensity region. The presence of deoxyhemoglobin is a strong indicator of the presence of cancer [28]. The encircled region with the malignant tissue in the PA image at 760 nm wavelength corresponds to the higher pixel intensity [28]. Previous works required to extract the encircled region of interest corresponding to cancer and non-cancer region by the co-registration of the histopathological slide, photograph-based image, and PA-based image.

Dogra et al. [28, 29] implemented the logistic function for cancer tissue detection using manually extracted features using the pixel-based samples. The handcrafted features were extracted by the pathologist and the expert in the field of PA imaging. Dogra et al. reported the accuracy of 0.9 on the prostate pixel-based samples [28] and accuracy of 0.83 on the thyroid pixel-based samples [29]. The pixel-based samples were extracted from normal, benign, and cancer regions from the PA image (the encircled region in PA image (8.2, (c))). The encircled region corresponding to the normal, benign, and cancer in PA image was generated by the co-registration of the histopathological slide and photograph image with the PA image (photograph (figure 8.2, (a)), histopathology slide (figure 8.2, (b)), and the two dimensional C-scan PA image (figure 8.2, (c))). The discriminant analysis was performed on PA image data that fell within the selected regions, not the entire image. All methods [28, 29] involved the manual, labor-intensive work to co-register the histopathological slide, and photograph to the PA image to extract three regions of interest such as normal, benign, and cancer. Moreover, the process was time-consuming to extract the optimal set of handcrafted features. That demanded the involvement of the expert with domain knowledge such as pathologist and image scientist [28, 29].

Dogra et al. [28] took one-slice C-scan out of 200-slice C-scan PA image (figure 8.3) at each wavelength and concatenated the five C-scans at five wavelengths to make five-channel C-scan images [28]. The five-channel C-scan image was then converted to 4-channel chromophore image (oxyhemoglobin (8.2, (e)), deoxyhemoglobin (8.2, (f)), lipid, and water). The set of pixels from the region of interest (normal and

malignant set of pixels from the encircled region) were collected and averaged to generate one pixel with four-channel chromophore (oxyhemoglobin, deoxyhemoglobin, water, and fat) image. Furthermore, only the pixel amplitude-based features were used, thus ignoring any possible (2D and 3D) spatially discriminant features. 2D spatial and 3D volumetric information is useful when the cancer tissue structure expands in all directions [48, 79].
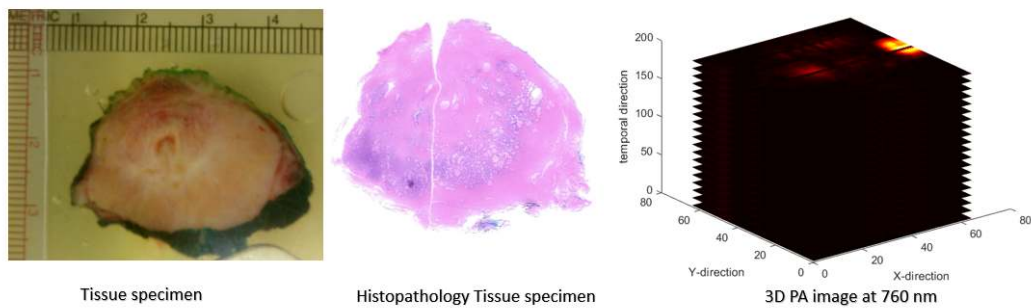
Figure 8.3: Histopathology specimen with the metric scale and 3D PA image of the specimen of size $64 \times 64 \times 200$, where first two dimension corresponds to the C-scan image and third dimension corresponds to the time-gated C-scan at different depth of the tissue specimen.

The machine learning classifiers used in the previous methods were trained, validated, and tested on the MPA pixel-based dataset, not in the whole MPA image dataset (Figure 8.4). These processes consisted of the extraction of the pixel-based MPA samples from the encircled region using the co-registration of the photograph, histopathological slide (suggested by the pathologist), and C-scan PA image. The process was labor-intensive and time-consuming, and yet the classifier is not fully

automated during testing.

This work was trained and evaluated on the whole MPA image, and hence, it removes the labor-intensive work required to extract the pixel-based samples using co-registration. That makes the current method is capable of detecting cancer tissue automatically [50, 51]. The previous studies using machine learning algorithms [28, 29] used the pixel-based handcrafted features such as pixel intensity but ignored the spatial and volumetric distribution of chromophores (blood). These work fail to extract spatial and volumetric information (figure (8.4)), which becomes essential when cancer tissue extends in all directions [48]. The process can be automated with the use of deep learning architecture consisted of convolutional neural networks (CNN)s.

Currently CNNs have been successfully used for a wide range of imaging tasks including image classification [9, 48], object detection [86] and semantic segmentation [14], and radiological text classification [49] in medical imaging. CNNs are popular because of their ability to extract the discriminant features at multiple levels of abstraction automatically without prior knowledge [97].

Jnawali et al [50] implemented 2D CNN using inception-resnet-v2 [96] for cancer tissue detection the spatial distribution of the chromophores with the AUC of 0.73. The method fails to extract the volumetric information (figure (8.4)). As the cancer tissue spreads in all directions, it is useful to implement the deep 3D CNN [48]. Jnawali et al. [51] improved the performance of cancer tissue detection with the use of the 3D CNN [50] with the AUC of 0.85. The improvement is due to that the
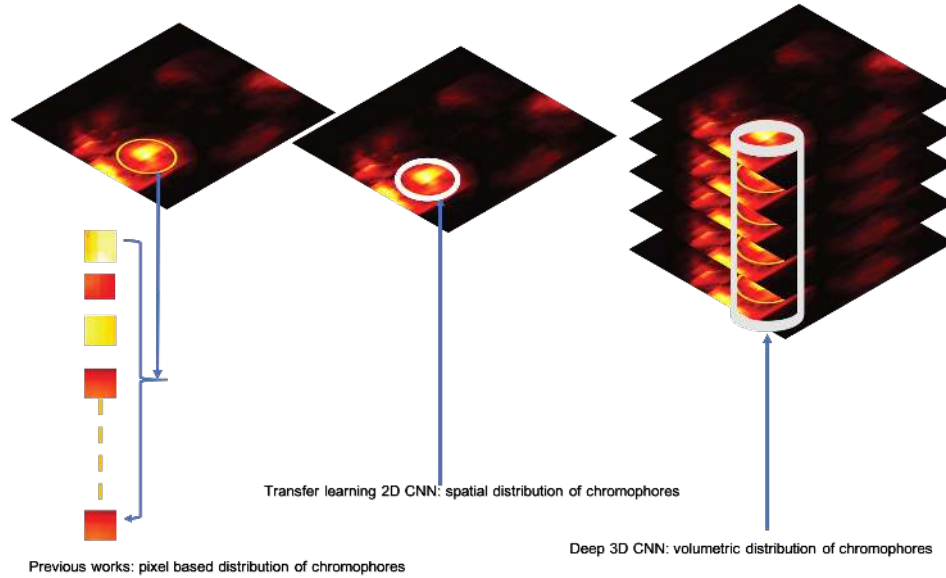
Figure 8.4: The figure shows the type of samples used to train, validate and, test the performance of the classifiers described in the previous studies (A) [28, 29], (B) [50], (C) [51] and the current study. The left figure shows the pixel-based samples used in the previous studies [28, 29]. The middle figure shows the concatenated one-channel C-scan PA image at three wavelengths samples used in the transfer learning network. The right figure shows the concatenated 21-channel C-scan PA image at five wavelengths samples used in the two 3D CNNs [51] and this paper. The left figure shows the requirement of the co-registration of PA image with the histopathological slide and the photograph of tissue specimen to extract the encircled region of interest during a time for a cancer tissue detection. The method used in the previous studies still requires manual work for co-registration during test time. This process becomes automated with the use of an entire MPA image during the test time for the cancer tissue detection.

deep 3D CNN is capable of automated extraction of volumetric blood concentration [48, 51] with the 3D convolutional kernel cube, which reduces cost, time of analysis and improves predictive performance. These two methods reduced the very time consuming manual work of feature extraction and co-registration [28, 29] and made

the detector complete automatic.

Jnawali et al. [51] implemented the deep 3D CNN with seven layers on the prostate MPA image with the AUC of 0.72, given the AUC of 0.85 on the thyroid test dataset, where author trained the deep 3D CNN on the thyroid MPA image [51]. Jnawali et al. [51] reported that there are common dominant discriminant features extracted by a PAI for cancer detection in thyroid and prostate. With this assumption, we created the train, validation, and test dataset with mixing the thyroid and prostate MPA image datasets with more balanced samples to train with deeper 3D CNN than the previous work [51]. A deep learning network generally improves with more samples [48], and the depth of the network (helps to extract more number of features with the use of more CNN filters) [38]. In this paper, the deep 3D CNN with more number of hidden layers is introduced for cancer detection and cancer region localization and compared to the deep 3D CNN with seven layers [51].

8.2   Ex-vivo PA dataset acquisition

Because we used a unique setup to acquire PA signal data set from tissue specimens, it is important to describe the method and the resulting data structure in this section briefly. The method, the details of which can be found in [103], is illustrated in figure (8.5). An expanded beam of 10 ns pulse of NIR laser light with delivered pulse energy below 20 mJ/cm^2 and pulse repetition rate of 10 Hz was used to expose the tissue specimen immersed in a water-filled medium [47]. PA signals in the form of a short

pulse of US generated from absorbers everywhere in the exposed 3D tissue volume propagate towards a specially designed acoustic lens of focal length f. The lens enables the simultaneous focusing of all the waves on the other side of the lens. If the center of the tissue is kept at a distance $2f$ from the lens, then a 32 element linear array of US transducers can be placed at $2f$ distance on the other side to detect the focused PA signals at 32 different pixel locations in the image plane for each laser firing. These US time signals, referred to as A-line signals, were amplified and then digitized at 30 MHz on 32 independent channels simultaneously. A-line signals were envelope detected in order to keep only the slowly varying nonzero signal values. The linear array was scanned in the image plane with repeated laser firing to collect PA signals over the entire image plane. The spatial resolution achieved by this system was around 1.3 mm [104]. PA signals from different depth planes along the lens axis arrive at the image plane at different arrival times due to the finite propagation speed of the US in water. By taking time slices on all the A-line signals, one can generate 2D C-scan PA images that correspond to different depth planes in the tissue as in figure (8.5). We were able to pick a time gate of 200 sample width, indicated by t_1 and t_1 in figure (8.5), that included all PA signals coming from the entire 3 mm thickness of every tissue specimen. The scanning in the image plane was typically done over a 40×40 pixels with a pixel size of 1 mm by 1 mm. The typical size of the 3D data for CNN study was $40 \times 40 \times 200$.

Every tissue specimen was imaged at five different NIR wavelengths, 760 nm, 800 nm, 850 nm, 930 nm, and 970 nm, generating 3D PA image dataset of typical

size $40 \times 40 \times 200$ for each. Patient consent was taken before prostatectomy and thyroidectomy following the University of Rochester medical center internal review board (IRB) protocol. One slice from the freshly excised gland was photoacoustically imaged with our set up within 15 to 30 minutes of surgery and then returned to the pathology department for further processing. Histology slide image and interpretation of that particular specimen was given to us later to be used as ground truth.
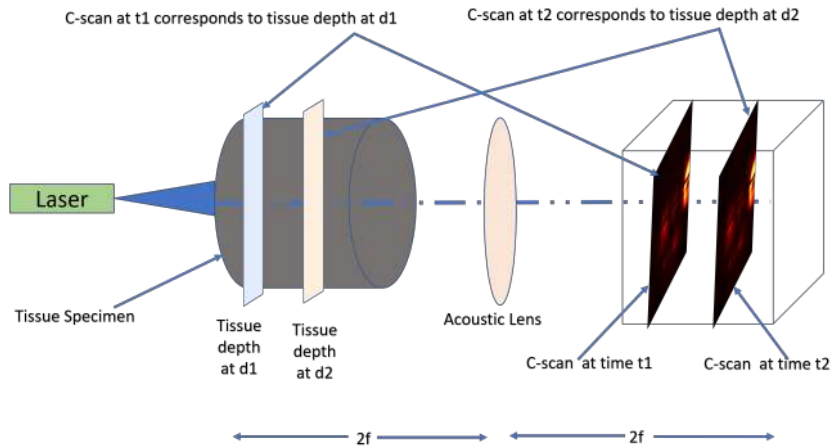
Figure 8.5: Schematic of the PA signal data acquisition setup. An acoustic lens was used to focus the photo-acoustically generated US waves and detect it by a linear US transducer array (not shown in figure). The C-scan image at the given depth is simply the magnitude of the raster scanned A-line signals at the given time for example t_1 (the time information was encoded to depth using $c = d \times t$, where c is the velocity of sound, d is the distance, and t is the time travel.) [103].

There is a trade-off between image quality and real-time imaging based on the transducer's number of elements. Furthermore, there are numerous techniques reported for the PA image reconstruction from the recorded PA signals by the trans-

ducer. There is a serious mathematical difficulty to use the Radon transform equation for an exact PA image reconstruction [116]. The algorithms namely back-projection [41], Fourier based [116], deconvolution based [113], iterative [22] were implemented for PA image reconstruction [116]. Rao et al. [76, 77] proposed the acoustic lens, which helps to circumvent the process of the PA image reconstruction [77]. The use of the acoustic lens makes the PA image acquisition process fast, robust, and less error-prone [77, 103].

Table 8.1: Normal and malignant tissue distribution

Tissue Specimen	Thyroid	Prostate	Total
Normal	91	17	108
Cancer	17	15	28
Total	108	28	136

8.3   Method: Deep 3D CNN with 11 layers

This section describes the step involving data preparation and the details of the algorithm for cancer tissue detection. This paper introduced the deeper deep 3D CNN with 11 layers than the deep 3D CNN with seven layers [51] and proved to be the best model to the date, to the best of our knowledge, compared with models [28, 29, 50, 51].

Data Preparation

Due to variability in data acquisition, the size of the data cube was not constant; it ranged from $20 \times 64 \times 200$ to $64 \times 64 \times 200$. The first two dimensions correspond to

the C-scan image along X, and Y direction in figure (8.6) and the third dimension corresponds to time samples along with the array of A-line signal (time-gated C-scan [104]). The third dimension reduced to 21 slices from 200 slices of C-scan PA image by taking the maximum of the sum of pixel intensity of C-scan and its ten neighbors. Equalization in the X and Y direction was done by increasing the dimension to 64 by 64 with bilinear interpolation, as shown in (8.6). The complete MPA dataset for each specimen was prepared by concatenating the five image cubes corresponding to wavelengths 760 nm, 800 nm, 850 nm, 930 nm, and 970 nm respectively, in that precise order (figure 7.2). The reason for choosing these wavelengths was because deoxyhemoglobin, oxyhemoglobin, fat, and water show peak absorption coefficient at 760 nm, 800 nm, 850 nm, 930 nm, and 970 respectively and the oxyhemoglobin and deoxyhemoglobin corresponds to same absorption coefficient at 800 nm [28].

Table 8.2: Training, validation, test, and augmented training dataset distribution.

Training	Validation	Testing	Augmented training
110	10	16	6,200

Algorithm

The proposed architecture consisted of four 3D convolutional layers, four 3D max-pooling layers, two fully connected layers, and a classifier as in figure (8.7). The network was trained up to 10 epochs with a batch size of 80 while saving the model corresponding to the lowest classification error using TensorFlow [2]. The network was initialized with the He initialization parameter that makes back-propagation
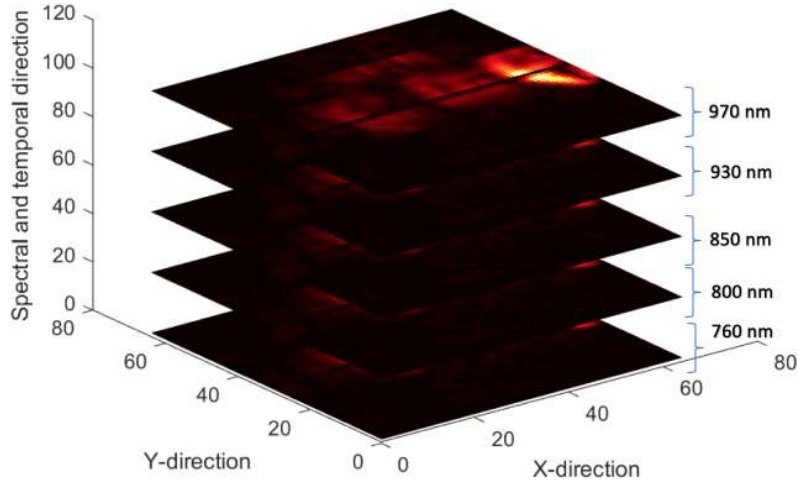
Figure 8.6: 105-channel (5 wavelength $\times$ 21-slice of C-scans PA image) of MPA dataset was prepared by stacking five 21-slice of C-scans PA image cube corresponding to five wavelengths, namely 760 nm, 800 nm, 850 nm, 930 nm, and 970 nm respectively. The 3D PA image cube at each wavelength is responsible for extracting the volumetric distribution of the chromophores (oxyhemoglobin, deoxyhemoglobin, lipid, and fat), and the 3D PA image cube at five wavelengths is responsible for imaging optically activated chromophores. For example, The deoxyhemoglobin is optically active at 760 nm and is responsible for a brighter pixel distribution in the PA image (figure 8.2 (c))

more efficient[38]. The backpropagation algorithm was used to minimize the loss function (A.1). The batch normalization [44] was applied after each convolutional layer and fully connected layer for faster training. The batch normalization also acts as a regularizer to prevent an overfitting [44]. A dropout of 0.5 was applied after each fully connected layer that acts as regularizer [93] to prevent the overfitting. Rectified linear unit (ReLu) activation function was applied after each convolutional layer and fully connected layer for efficient backpropagation by reducing the gradient vanishing problem [59]. The network's model parameters were updated with the stochastic

gradient descent [57] during the backpropagation. The network was trained with a learning rate initialized at 0.01 and decayed exponentially in each epoch by 0.1 as in figure (8.8). The hyperparameter set {number of convolutional layers, size of convolutional filters, stride size, max-pooling layers, number of fully connected layers, number of nodes in the fully connected layer, activation function, dropout, and learning rate} were chosen using a grid search started from the deep 3D network described in the model 1 [48]. The first, second, third, and fourth convolutional layers (colored in black) consisted of 8, 16, 24, 32 filters respectively with kernel size $3 \times 3 \times 3$. The max-pooling layer (colored in gray) was used to reduce the number of features from the convolutional layer that helps to keep the deep learning network invariant to translation and small rotation [15]. The pooling size of the max-pooling layer was set at $2 \times 2 \times 2$. The two fully connected layers with 512 and 128 nodes respectively added to the network after the max-pooling layer. The detail architecture is shown in figure (8.7) with its hyperparameters in table (8.3). Finally, a softmax function was applied to the network for the binary cancer tissue detection.

Multiclass cross entropy error function is given by the equation [15]

$$L = -\sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk} ln \; y_{nk} \qquad (8.1)$$

$$y_k(\phi) = \frac{exp(a_k)}{\sum_j exp(a_j)} \qquad (8.2)$$

where $y_k(\phi)$ represents a softmax function, K is number of class, t_k is the class
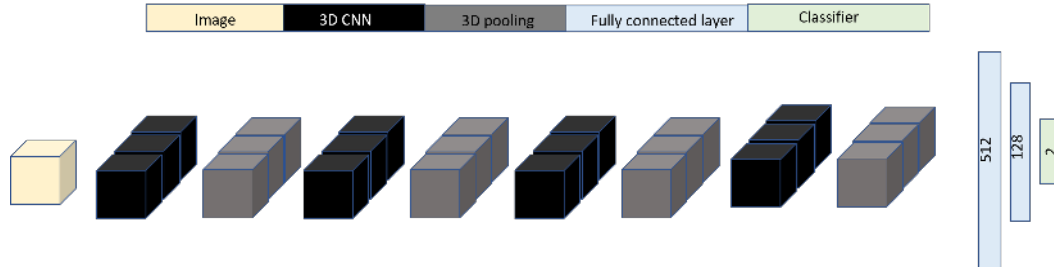
Figure 8.7: The detail architecture of the network used in the experiment 3 for cancer tissue detection which is deeper than that used in the experiment 2
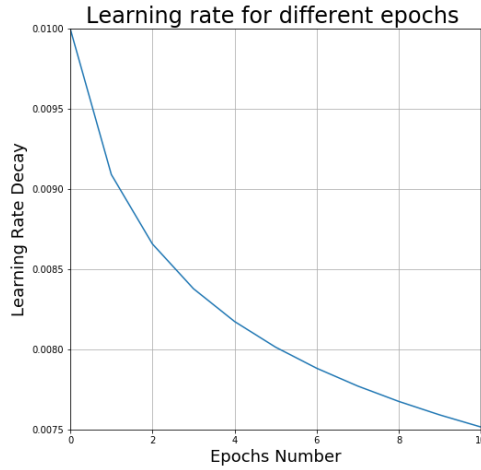
Figure 8.8: learning rate decay in each epochs

target vector of 1-of-K coding scheme, a_k is given by $W_k^T \phi$, y_{nk} is given by $y_k(\phi_n)$ [15].

The 136 thyroid and prostate MPA image datasets were used to train, validate, and test the proposed deep 3D neural network and the network. The MPA image dataset was divided into the train (120 specimens) and test data (16 specimens) in the ratio of 0.9:0.1 respectively. The train data was further divided into the 0.9:0.1

Hyperparameters	Input	Output
Convolutional layer 1	$64 \times 64 \times 105$	$8 \times 64 \times 64 \times 105$
Max pooling layer 1	$8 \times 64 \times 64 \times 105$	$8 \times 32 \times 32 \times 52$
Convolutional layer 2	$8 \times 32 \times 32 \times 52$	$16 \times 32 \times 32 \times 52$
Max pooling layer 2	$16 \times 32 \times 32 \times 52$	$16 \times 16 \times 16 \times 26$
Convolutional layer 3	$16 \times 16 \times 16 \times 52$	$24 \times 16 \times 16 \times 26$
Max pooling layer 3	$24 \times 16 \times 16 \times 26$	$24 \times 8 \times 8 \times 13$
Convolutional layer 4	$24 \times 8 \times 8 \times 13$	$32 \times 8 \times 8 \times 13$
Max pooling layer 4	$32 \times 8 \times 8 \times 13$	$32 \times 4 \times 4 \times 6$
Fully connected layer 1	$32 \times 4 \times 4 \times 6$	512
Fully connected layer 2	512	128
Classifier layer 1	128	2

Table 8.3: Hyperparameters of the deep 3D CNN with 11 layers

for the training (110 specimens), and validation (10 specimens) MPA image dataset respectively. The network was trained with five-fold cross-validation [15]. The cross-validation method, while training a classifier is useful to minimize the variance in the model [33]. The network was trained and validated with a mixture of prostate and thyroid MPA dataset. The number of samples was increased using image augmentation [59] before training the network. This increases the samples and is useful when the dataset is limited. The MPA dataset was augmented by random pixel shift from $(1, 10)$ pixels, random pixel rotation from $(-40, 40)$ degrees, random noise from $(5, 15)$ pixel intensity, random cropping, random warping, vertical flipping, and horizontal flipping. The figure shows three methods of augmentation as in figure (8.9) [16]. The cancer MPA dataset was augmented almost five times more than the normal MPA dataset to make normal cancer more balanced. The distribution of data was remained imbalanced after augmentation. In addition to augmentation, a class-weight parameter [118] was introduced while training the network. The class

weight function looks at the distribution of positive and negative samples and produces weights to equally penalize under or over-represented classes to minimize the problem due to imbalance class distribution [49]. Finally, the network for the cancer tissue detection was evaluated by generating an AUC of ROC curve, a metric of choice when the dataset becomes imbalance [14] and provides a more degree of freedom for threshold choice [14]. The validation dataset was used to optimize the hyperparameters of the network, which helps to improve the classifier's predictive performance. A grid search method was used to optimize the hyperparameter for the proposed model. The set of hyperparameters used for the model is tabulated in table (8.3).
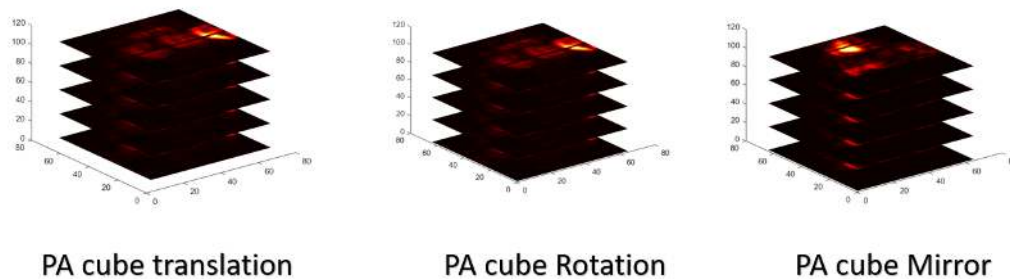
Figure 8.9: The MPA dataset was augmented by random pixel shift from $(1, 10)$ pixels, random pixel rotation from $(-40, 40)$ degrees, random noise from $(5, 15)$ pixel intensity, random cropping, random warping, vertical flipping and horizontal flipping. The figure shows three methods of augmentation.

8.3.1   Results

Figure (8.10) shows the activation map of the deep 3D CNN implemented in this paper. The activation map for the cancer tissue shows a wider distribution of the

mass than that for the normal tissue, a key feature to detect cancer [69]. This could be due to a higher blood distribution around the malignant tissue [11].
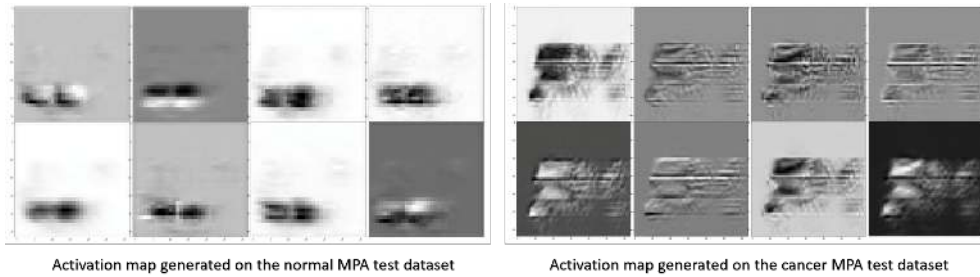
Figure 8.10: Activation map generated by the deep 3D CNN on the normal and cancer MPA test image. Xu et al. [116] reported that the malignant tissue is more likely to be in irregular shape and normal tissue is more likely to be in rounded shape.
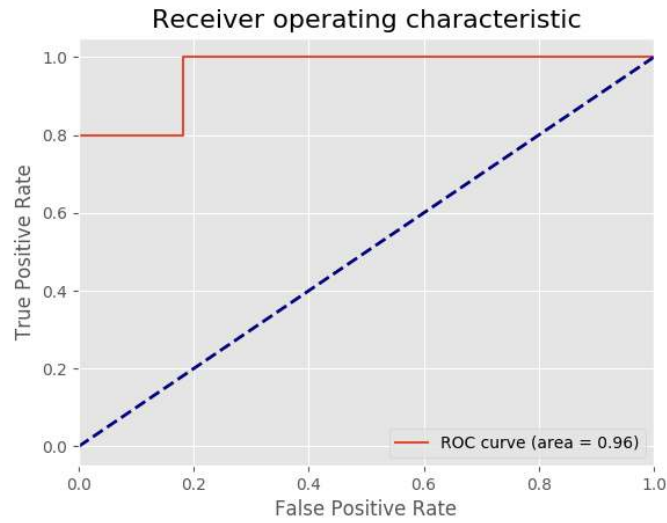
Figure 8.11: ROC curve generated by the model on the test MPA image.

The proposed 3D CNN with 11 layers generated the AUC of 0.96 as in figure (8.11) higher than the previous works using transfer learning [50] and deep 3D CNN with seven layers [51]. This model shows the improvement over the earlier models described [50, 51] because of deeper architecture [39], and more samples [48]. The proposed deep 3D CNN with 11 layers is capable of extracting all spatial, volumetric, and spectral features with the AUC of 0.96. The accuracy improvement is expected with more number of samples [51]. The algorithm described in this paper reduces labor and time-intensive human involvement for image co-registration, and handcrafted feature extraction [28] for cancer tissue detection. The confusion matrix and precision, recall, f_1 score are shown in the tables below. Table (8.6) shows the metric and its value used to measure the classification performance.

		Predicted		
	Case	Negative	Positive	Total
Ground Truth	Negative	11	0	11
	Positive	1	4	5

Table 8.4: Confusion Matrix

	Precision	Recall	$f1$-Score	Support
Normal	0.92	1.0	0.96	11
Cancer	1.0	0.8	0.89	5
Avg	0.94	0.94	0.94	16

Table 8.5: Precision, recall, f_1- score and support

Algorithms	Feature used	Dataset	Metric	Scalar value	Author
Logistic function	Handcrafted	Prostate	Accuracy	0.9	Dogra et al. [28]
Logistic function	Handcrafted	Thyroid	Accuracy	0.83	Dogra et al. [29]
2D Transfer learning (inception-resnet-v2)	2D image based	Thyroid	AUC	0.72	Jnawali et al. [50]
deep 3D CNN (seven layers)	3D image based	Thyroid/Prostate	AUC	0.85/0.72	Jnawali et al. [51]
deep 3D CNN (11 layers)	3D image based	Thyroid and Prostate	AUC	**0.96**	This paper

Table 8.6: The proposed model is compared with the previous models [28, 29, 50, 51]

8.4   Automatic Localization of the Cancer Tissue Region

The aim of this section is not only to detect cancer in the given MPA image dataset but also to localize it with additional confidence. There is numerous approach described in the literature to visualize the region where CNN is giving more importance.

Selvaraju et al. [84] proposed the gradient-weighted class activation mapping for the visualization of the areas where deep learning network is looking into for a given object. The author used the VGG-network to detect the region where the neural network is looking. In this thesis, we implemented the grad-cam algorithm using the proposed deep 3D CNN with 11 layers for the cancer region localization automatically with the promising result as in figure (8.10).
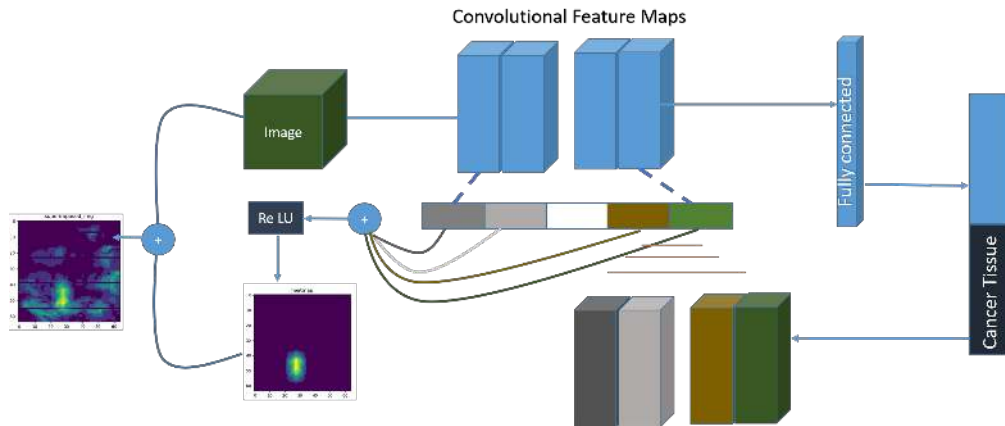
Figure 8.12: The automatic localization using grad-cam is very useful not only to detect the object with localization but also to shed light on how the given neural network architecture is working.

8.5 Discussion and Conclusion

CNNs generally require a large dataset for training. PA imaging being a new medical imaging modality clinical data is not yet available. The PA dataset used in this work is, to the best of our knowledge, is the largest available human ex-vivo dataset on thyroid and prostate tissue specimens. Jnawali et al. [51] hypothesized that the underlying PA discriminant features for cancer and non-cancer have a common trend in thyroid and prostate cancer detection, and hence they trained on thyroid and applied the model to the prostate MPA image. Thyroid was chosen for training because this set had a more substantial number of specimens (table (8.1)). Given the AUC of 0.85 on the thyroid dataset, the AUC of 0.72 on the prostate dataset using the deep 3D CNN with seven layers (trained and evaluated on the thyroid MPA image samples) is encouraging and supports the hypothesis [51].

In this paper, the next step was to mix the dataset with the thyroid and prostate, given the result from previous work [51]. The number of samples to train the network for this paper was more significant than previous work [51] and hence decided to develop a deeper 3D CNN. A deeper CNN is more powerful in predictive performance compared to a shallower one [38, 96]. The AUC of 0.96 is much better compared to the previous two deep learning methods. There could be multiple reasons for this improvement. The first reason is the larger number of the dataset, with similar deterministic features, the second reason is the dataset becoming more balanced between cancer and normal (table (8.1)), and the third region is the more number of CNN filters (more number of features) [96]. This suggests that in situations where

the number of a dataset is small, this type of approach is reasonable to train CNNs adequately.

This work is different from the previous works using machine learning algorithms [28, 29]. There, with the pathology ground truth, the data from cancer and normal regions were isolated from the encircled area of each specimen and then used for discriminant analysis — this required extensive expert human involvement of pathologists and image processing personnel. The cancer region marked by the pathologist on the histopathology slide had to be co-registered and transferred over to the PA images [28]. The process is still not automated because of the extraction of the pixel-based samples during test time. The use of the CNN model is able to circumvent this labor-intensive human involvement.

One plausible advantage of our current work using the deep 3D CNN is that the model is able to learn from the spatial, volumetric, and spectral features, all at once. This model may have immediate application in cancer screening of the numerous sliced specimens that result from thyroidectomy and prostatectomy. Since the instrument that was used to capture the ex-vivo PA images is now being developed for in-vivo use, this model may also prove to be a starting point for in-vivo PA image analysis for cancer diagnosis. This model is also capable of localizing the cancer region in the given MPA image automatically. This method not only detects cancer but also locate the cancer region in the given MPA image.

8.6   Future Work

There are two options for improving the predictive performance of CNN for cancer tissue detection using the MPA dataset. The first option is to collect more similar specimens dataset to be included in the model building process because the performance is known to increase with the number of dataset [48]. The second option is to modify the inception-resnet-v2 [96] to 3D inception-resnet-v2 and then used with the existing MPA dataset; this is because there is some evidence that 2D inception-resnet-v2 [50] gives a promising result and extension to 3D will very likely result in further improvement.

8.7   Acknowledgments

Chapter 9

Conclusions and Future Works

Photoacoustic imaging (PAI) is a new medical imaging modality [104] that is currently making a transition from bench to bedside, both in terms of technology [104] and clinical applications [28]. In this thesis, we discussed the potential of PAI modality for cancer tissue diagnosis in the near future. The 3D multispectral PAI is capable of extracting the optical signature and volumetric distribution of the optical biomarkers for cancer diagnosis such as oxyhemoglobin, deoxyhemoglobin, lipid, and water [28]. These biomarkers have been proved to be the key attributes for the early cancer diagnosis using MPA image [28, 61]. CT or MRI imaging is not capable of detecting cancer tissue in the early stage and is capable of detecting cancer tissue only when the diameter grows to 1 cm in size [61]. It has been already proven that early detection of cancer tissue increases the chance of cure [28] and PAI offers to detect cancer in the early stages [61]. PA image is safe due to their non-ionizing ra-

diation properties [104] unlike CT image. This makes PAI is capable of imaging soft tissue noninvasively and enhances its clinical feasibility in the future. Its low-cost feature too a great advantage over other imaging techniques like MRI or CT [34].

Dogra et al. [28] and Rajanna et al. [75] proposed machine learning algorithms, namely logistic classifier, and two-layer neural network respectively for cancer tissue detection with manually extracted features based on the MPA pixel-based samples. Pixel-Based cancer and normal samples were extracted from the region of interest. The pathologist provided the histopathological slide with the marking of normal and cancer as the region of interest. This required to co-register histopathological-based image and photograph-based image and transfer over to the PA images for the extraction of the region of interests - normal and cancer region. The mathematically extracted handcrafted pixel-based features were used to build the classifiers. Their classifiers were trained, validated, and tested on the pixel-based samples obtained from the region of interest of the PA image. There is still manual involvement of the pathologist to mark the region of interest during test time as their classifiers still require to co-register the histopathological slide to the PA image to extract pixel-based samples corresponding to the normal and cancer class during testing. The models still require the labor-intensive and time-consuming co-registration work during the test. These models still demand the marking of the label (region of interest) by the pathologist during testing, that is generally not the case in the machine learning approach. This is because all models should be automated during testing at least [15]. In the machine learning approach, the feature set is extracted by the expert with the

domain knowledge during training. However, the extraction of features can be automated with the use of deep learning architectures, especially convolutional neural networks (CNNs). The use of the CNN model is able to circumvent this extraction of handcrafted features for cancer tissue detection. The current study detects cancer tissue using the MPA image automatically using deep learning algorithms and is different from the previous studies, which still requires manual work at least for the co-registration. The current method solves the problem of co-registration by using entire C-scan image where involvement of pathologist during testing is none unlike previous studies [28, 75].

The three models were implemented for cancer tissue detection starting from the transfer learning using inception-resnet-v2 [96], deep 3D CNN with seven layers [51] and deep 3D CNN with 11 layers. All three models take the entire C-scan image as the input and hence no involvement of the pathologist during testing, which is not the case in the previous studies [28, 75]. The transfer learning network was implemented to evaluate the predictive performance of a deep learning algorithm to detect cancer tissue. The transfer learning network using inception-resnet-v2 is quick to implement and train. The number of input channels used in the inception-resnet-v2 is three [96]. So the three-channel MPA image input of the model was prepared by concatenating the one-slice of C-scan image with the sum of maximum pixel intensity at three wavelengths, namely 760 nm, 800 nm, and 850 nm. The model was able to detect cancer with the AUC of 0.73 on the thyroid three-channel MPA image [50]. Since the model was unable to use data consisting of all five

wavelengths and the entire depth of the tissue specimen, we chose to develop a deep 3D CNN model. The input of the model was prepared by concatenating the 21-slice of C-scan image at five wavelengths, namely 760 nm, 800 nm, 850 nm, 930 nm, and 970 nm. The input dataset for the deep 3D CNN architecture was able to contain the volumetric distribution of the chromophore at five wavelengths. The deep 3D CNN with seven layers was trained and evaluated using the thyroid 105 (5 wavelengths × 21)-channel MPA image dataset. The model was also evaluated on the prostate MPA test dataset to validate the hypothesis that PA imaging modality is capable of extracting the characteristic feature of chromophores in two tissue specimens, namely thyroid, and prostate [51]. The model was able to generate the AUC of 0.85 on the thyroid MPA image dataset and 0.72 on the prostate MPA image dataset. We mixed the thyroid and prostate MPA 105-channel MPA image dataset together to increase the sample size. The decision of mixing two datasets together was based on our previous results, which suggest that there is a similar trend in classifying cancer form normal with prostate and thyroid tissue [28]. The performance of deep learning algorithms generally improves with the number of samples [48] and helps to make deeper CNNS [97]. The deep CNNs are a generally good choice for the robust predictive performance but limited to the case with a smaller number of samples [15]. Finally, the deep 3D CNN with 11 layers was trained with a mixture of thyroid and prostate 105-channel MPA image dataset with a higher number of samples compared to the previous studies. The model (deep 3D CNN with 11 layers) was also used to automatically localize the cancer tissue region in the given MPA image test dataset

using the grad-cam algorithm [84].

The model (deep 3D CNN with 11 layers) is not only the improvement over the previous works using machine learning [28, 29] and deep learning algorithms [50, 51] but also capable of detecting and localizing the prostate and thyroid cancer tissue at once. One plausible advantage of our current model using the deep 3D CNN with 11 layers is to learn spatial 2D and volumetric 3D features of chromophores all at once for both thyroid and prostate specimens. The model was also implemented to localize the cancer region in the given MPA dataset for additional confidence where the deep 3D CNN is looking into during cancer tissue detection. The recommendation of the region of interest (cancer) also could save the time of pathologists for early screening. This model may have immediate application in cancer screening of the numerous sliced specimens that result from thyroidectomy and prostatectomy. Since the instrument that was used to capture the ex-vivo PA images is now being developed for in-vivo use, this model may also prove to be a starting point for in-vivo PA image analysis for cancer diagnosis.

Appendices

Appendix A

Brain Hemorrhage Classification

This work is taken from the paper published in SPIE Medical Imaging, 2018.

[Kamal Jnawali, Mohammad R Arbabshirani, Navalgund Rao, and Alpen A Patel. Deep 3d convolution neural network for CT brain hemorrhage classification. In Medical Imaging 2018: Computer-Aided Diagnosis, volume 10575, page 105751C. International Society for Optics and Photonics, 2018.] [48]

Abstract

Intracranial hemorrhage is a critical conditional with high mortality rate that is typically diagnosed based on head computer tomography (CT) images. Deep learning algorithms, in particular convolution neural networks (CNN), are becoming the methodology of choice in medical image analysis for a variety of applications such

as computer-aided diagnosis and segmentation. In this study, we propose a fully automated deep learning framework which learns to detect brain hemorrhage based on cross sectional CT images. The dataset for this work consists of 40,367 3D head CT studies (over 1.5 million 2D images) acquired retrospectively over a decade from multiple radiology facilities at Geisinger Health System. The proposed algorithm first extracts features using 3D CNN and then detects brain hemorrhage using the logistic function as the last layer of the network. Finally, we created an ensemble of three different 3D CNN architectures to improve the classification accuracy. The area under the curve (AUC) of the receiver operator characteristic (ROC) curve of the ensemble of three architectures was 0.87. The results are very promising considering the fact that the head CT studies were not controlled for slice thickness, scanner type, study protocol or any other settings. Moreover, the proposed algorithm reliably detected various types of hemorrhage within the skull. This work is one of the first applications of 3D CNN trained on a large dataset of cross sectional medical images for detection of a critical radiological condition.

A.1   INTRODUCTION

Deep neural network is machine learning framework with a wide range of applications from natural language processing [68] to medical image classification [86] and segmentation [8]. In particular, convolution neural networks (CNNs) have become a very popular method in the field of computer vision community. They have been used in variety of tasks including object detection, classification [59], [97] and localization. The power of CNN deep architecture is to automatically extract the

discriminant features at multiple levels of abstraction. The early few layers of CNN learn low-level image features and later layers learn high-level image features which are specific to the type of applications.

Before 2012, traditional computer vision techniques based on handcrafted features were popular for medical image analysis [65]. In 2012, Krizhevsky et al [59] introduced a deep convolution neural network (DCNN) which outperformed traditional methods by a significant margin in image classification and localization [65]. In such models, the features are learned and extracted automatically. A Convolution layer in a CNN detects certain local features from input images. A filter with a set of shared weights (called convolution kernel) is moved around the image to extract the features. Therefore, a convolution layer with k kernels detects k local features. Some convolution layers are followed by max pooling layer which reduces the feature size by selecting the maximum feature response. Max pooling helps with translational invariance [98] and to some extent rotational invariance. A neuron in the convolution neural network has very limited receptive filed with shared weights. Therefore, fully connected layers are typically used after convolutional layers (where each hidden neuron connects to all previous hidden neurons with individual weights). Finally, a logistic regression layer is used to get the desired output, in our case classification. The back-propagation algorithm is used to trained CNNs to minimize the cost function of the form [A.1]:

$$L = -\frac{1}{N} \sum_i^{|N|} \ln(p(y^i | X^i)) \qquad (A.1)$$

where N is the number of training images, y^i and X^i are i^{th} label and image respectively. The weights in the convolution layer l is updated using stochastic gradient descent:

$$W_l^{t+1} = W_l^t - \alpha(t)\frac{\partial \hat{L}}{\partial W_l} \tag{A.2}$$

where $\alpha(t)$ is a learning rate at the specific epoch and $\alpha(t+1)$. Learning rate at the following epoch, is given by the following equation:

$$\alpha(t+1) = \alpha(t) \times \frac{1}{(1 + \text{decay} \times \text{epoch(t)})} \tag{A.3}$$

where, the decay is given by the following equation:

$$\text{decay} = \frac{\alpha(t)}{\text{Total epoch}} \tag{A.4}$$

2D CNN is mostly used in natural RGB images to extract the spatial features only in two dimension [59] [97]. Though 2D CNN can be applied to the volumetric data set (such as cross sectional CT images), some researchers [65] have reported the benefit of 3D CNN to incorporate the volumetric information in medical images to get a higher accuracy. In our experiment, we extend 2D CNN to 3D CNN for the brain hemorrhage classification since the spatial information of the brain hemorrhage in CT image extends to three dimension. This work as an extension of our previous work using 2D DNN architecture [7] We modified the AlexNet type [59] and GoogLeNet type[97] architecture to 3D CNN structure with a fewer number of

hidden layers and network parameters.

A.2 Dataset and Data Processing

The head CTdataset was collected retrospectively from Geisinger Health System (GHS). It consists of 40,357 CT brain images with 30,001 negative cases (no hemorrhage) and 10,356 with a brain hemorrhage termed as a positive case. The total number of data set present in the training set are 34,848 with 26,383 as the negative case and 8465 as the positive case. The available CT image cube has the dimension ranging from $12 \times 512 \times 512$ to $150 \times 512 \times 512$. We reduce the size of the CT image cube to $24 \times 256 \times 256$ using cubic spline interpolation technique. One sample of CT is shown in figure [**??**]. The data set is highly unbalanced and it is a typical concern in the field of medical image analysis. We augmented the data set up to balanced 276,237 images by translation, rotation, and mirroring to reduce the issues of the imbalanced dataset. We used 2,000 studies for validation and 3,509 for testing the algorithm.

A.3 Method

In this study, we explore, evaluate and analyze the influence of various CNN architectures. Three type of 3D CNN architectures are proposed and are combined and averaged to make a single model. We explore two different 2D CNN architectures AlexNet [59] and GoogLeNet [97] and extend them to 3D CNN with numerous mod-
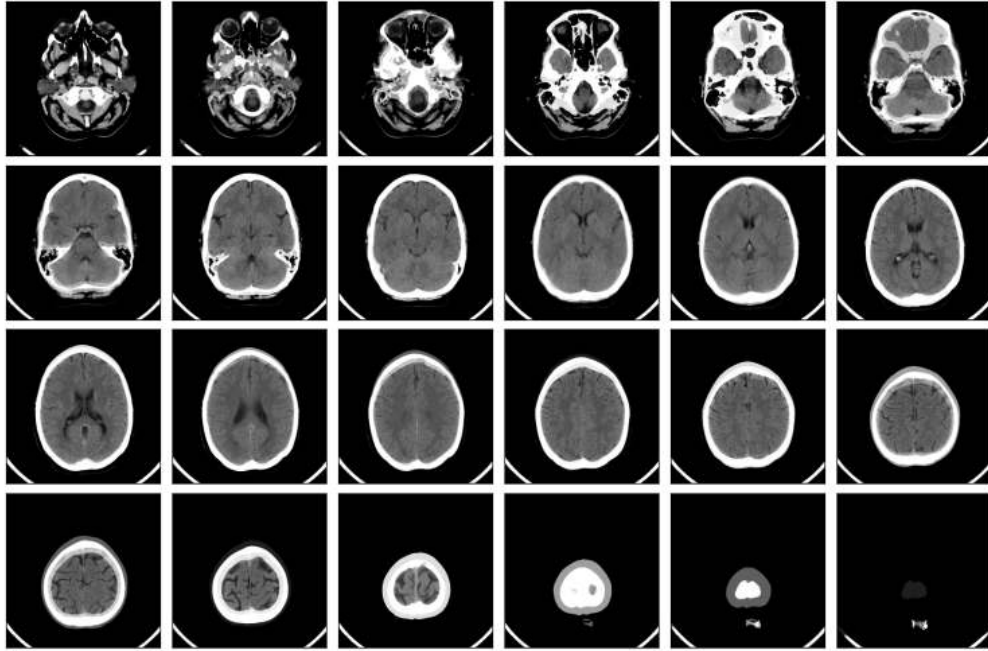
Figure A.1: One sample of CT image used in our experiment

ifications. Each architecture is modified to 3D CNN with a fewer number of deep layers and a fewer number of parameters to minimize the computational complexity in terms of space and time. Originally, two deep CNN network architectures, AlexNet [59] and GoogLeNet [97], are designed for the fixed image dimension of the $256 \times 256 \times 3$ with 60 million and 21 million parameters respectively. Those architectures are very expensive in terms of time and space complexity. We reduce the number of layers, filter size, stride and pooling parameters of the AlexNet type and GoogLeNet type architectures to (6,456,717 and 5,217,417) and 11,486,845 respectively to address the issue of space and time complexity in the proposed 3D

CNN.

A.4   Experiments and Results

A.4.1   Experiments

The three 3D CNN architectures were programmed in Tensorflow [**?**]. This networks were trained with augmented 276,237 data set for 10 epochs with a training batch size of 74 using a batch normalization technique [44] that accelerates the deep learning training by reducing the co-variance shift. The batch normalization which acts as a regularizer [44]was applied after each ReLu layer of the network. A ReLu activation function was applied after each layer of the networks except the last layer where a sigmoid activation function was applied. The ReLu activation function is extremely efficient to avoid the vanishing gradient problem during the back-propagation [59]. The networks were made more robust further with He initialization [38] that makes back-propagation more efficient. He initialization draws samples from a truncated normal distribution with mean of value zero and standard deviation of value $\sqrt{\frac{2}{N}}$, where N is the number of input units in the weight tensor [38]. The networks were trained with the learning rate equal to 0.01 and is decayed by a factor given by equation [A.3] for each epoch. The learning rate decay pattern is shown in figure [A.5]. Dropout was applied after each fully connected layer which acts as a regularizer [93] to overcome an over-fitting issue. Finally, the probability of the class is read from the output of the logistic regression layer, the last layer of the networks, for generating the ROC curves. The detailed information of the architectures are

described below.

A.4.2   First Architecture

This network had four CNN layers, two max pooling layers, two fully connected layers, and one output layer (figure [A.2]). Dropout [93] of value equals to 0.6 was applied after each fully connected layer. The total number of parameters for this network was 6,456,717.
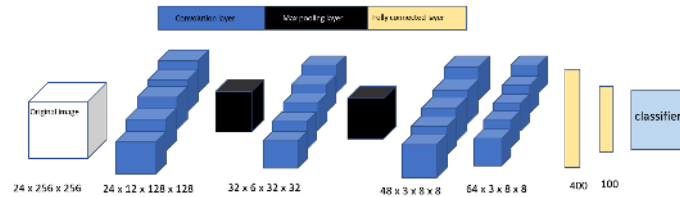
Figure A.2: First Architecture : Model 1

A.4.3   Second Architecture

This network had two CNN layers, two max pooling layers, one inception module [97] two fully connected layers, and one output layer (figure [A.3]). The inception module block consists of 1×1, 3×3, and 5×5 filters Dropout of value equals to 0.8 was applied after each fully connected layer. The total number of parameters for this network was 11,486,029.
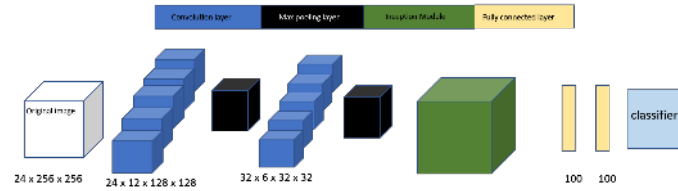
Figure A.3: Second Architecture : Model 2

A.4.4 Third Architecture

This network had four CNN layers, two max pooling layers, two fully connected layers, and one output layer (figure [A.4]). Dropout of value equals to 0.8 was applied after each fully connected layer. The total number of parameters for this network was 5,217,417.
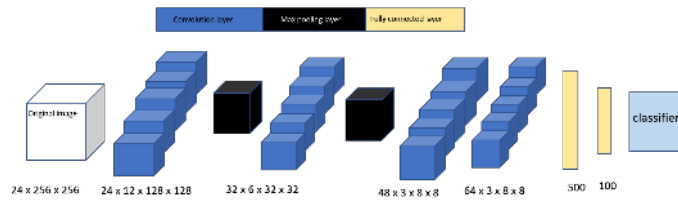
Figure A.4: Third Architecture : Model 3
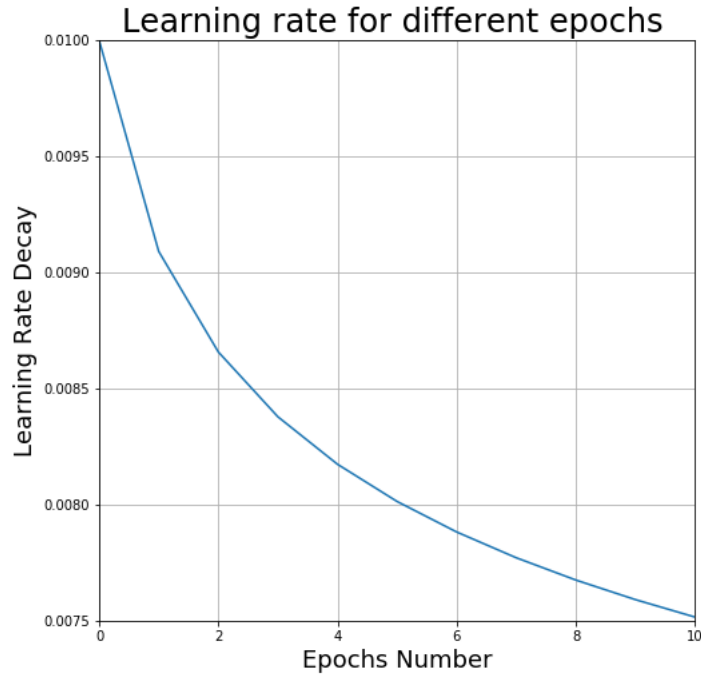
Figure A.5: Visualization of the rate of decay of the learning rate with each epoch

A.4.5 Evaluation and Result

The deep learning architectures were trained, validated and tested by splitting the dataset into three group, train-dataset, validation-dataset and test-dataset. The model is validated with the validation-dataset to pick the best choice of the weight matrix of the deep learning architecture that minimizes the classification error. The classification error minimization task in the deep learning architecture also demands a very careful choice of hyperparameters, for example, number of hidden layers,

number of hidden units in the fully connected layer, drop out, learning rate etc. We experimented with various value of drop out in the range of 0.0 to 1.0 and the value corresponding to the minimum classification error was selected.

The choice of metric to evaluate the performance of a classifier is crucial in the field of medical imaging where an unbalanced dataset is a typical concern. The number of the studies without a brain hemorrhage was significantly larger than that with a brain hemorrhage in this work. In contrast, a natural image dataset is very balanced and in such case the accuracy is a typical choice of metric to measure the performance of a classifier[59] [97].

The ROC curve is a popular metric to evaluate the performance of a classifier even for unbalanced datasets. For each architecture, the classifier performance was evaluated using the AUC of the ROC curve [17]. The AUCs for three different architectures at the threshold of 0.5 were 0.85, 0.85 and 0.86. The AUC of the ensemble of the three architectures at the threshold of 0.5 (illustrated in figure [A.6]) was 0.87.

At threshold of 0.5, the confusion matrix of the ensemble of three architecture is summarized in Table A.1. The table shows the distribution of the ground truth (actual) and predicted dataset.

The recall, precision and f_1- score of the ensemble of three architecture at threshold of 0.5 are shown in Table A.2.
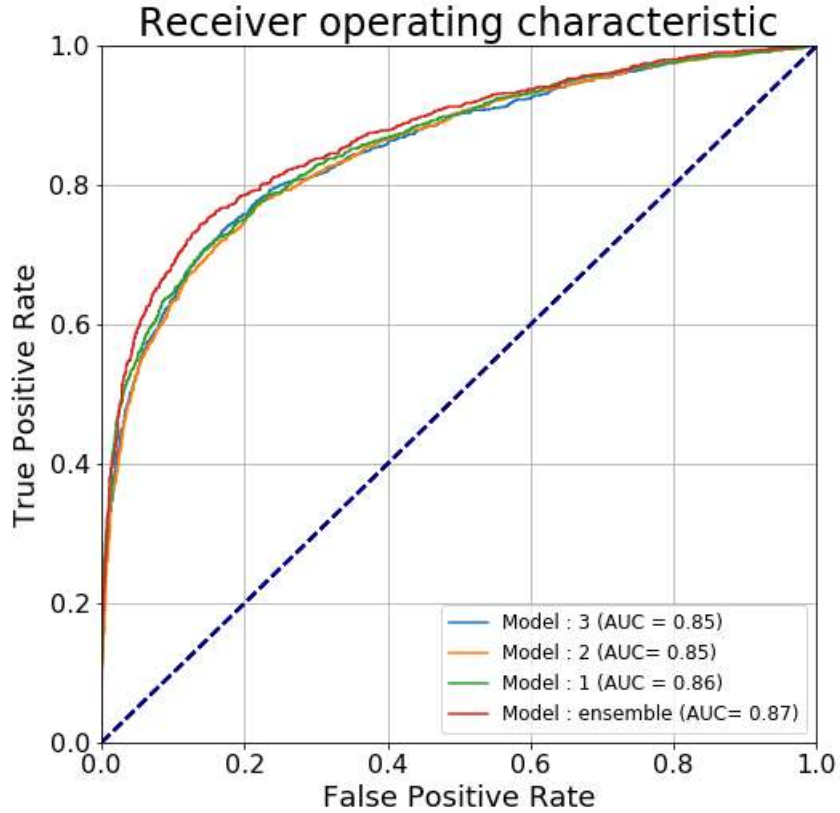
Figure A.6: Generated ROC curve based on the ensemble of three 3D architectures

A.5 Conclusion

In this work, we implemented three different 3D CNN architectures for detection of brain hemorrhage on head CT images. We augmented the imbalanced CT image data set to create a training dataset of approximately 250,000 samples. The ensemble

		Predicted		
	Case	Negative	Positive	Total
Ground Truth	Negative	1797	520	2317
	Positive	270	922	1192

Table A.1: Confusion Matrix

	Precision	Recall	f_1- score	Support
0	0.87	0.78	0.83	2317
1	0.64	0.77	0.70	1192
Avg/Total	0.80	0.77	0.78	3509

Table A.2: Precision, recall, f_1- score and support

of the three 3D architectures resulted in AUC of 0.87 which is very promising. To our knowledge, this is one of the first examples of 3D CNN applied to large head CT data set to detect a critical condition such as intracranial hemorrhage. The results may be a step toward wider adoption of deep learning in modern computer-aided diagnosis.

One future direction of this work is to explore automated technique to pick the optimal hyperparameters to minimize the classification error of the deep learning architectures. The random search [13] and bayesian optimization technique [92] are popular automated technique to obtain the optimal value of the hyperparameter of deep learning architectures. These two approaches greatly reduce the computational time of the deep learning architecture to find the optimal set of hyper parameter automatically.

A.5.1 Future work

One future direction of this work is to explore automated technique to pick the optimal hyper-parameters to minimize the classification error of the deep learning architectures. The random search and bayesian optimization technique. The random search [13] and bayesian optimization technique [92] are popular automated technique to obtain the optimal value of the hyperparameter of deep learning architectures. These two approaches greatly reduce the computational time of the deep learning architecture to find the optimal set of hyper parameter automatically.

Bibliography

[1] American Cancer Society. cancerstatisticscenter.cancer.org. 2018.

[2] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*, 2016.

[3] Agarwal, A and Huang, SW and O'donnell, M and Day, KC and Day, M and Kotov, N and Ashkenazi, Shai. Targeted gold nanorod contrast agent for prostate cancer detection by photoacoustic imaging. *Journal of applied physics*, 102(6):064701, 2007.

[4] Jose M Alvarez and Mathieu Salzmann. Learning the number of neurons in deep networks. In *Advances in Neural Information Processing Systems*, pages 2270–2278, 2016.

[5] Stephan Antholzer, Markus Haltmeier, and Johannes Schwab. Deep learning

for photoacoustic tomography from sparse data. *Inverse problems in science and engineering*, 27(7):987–1005, 2019.

[6] Maria-Luiza Antonie, Osmar R Zaiane, and Alexandru Coman. Associative classifiers for medical images. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 68–83. Springer, 2002.

[7] Fornwalt Brandon K. Mongelluzzo Gino J. Suever Jonathan D. Geise Brandon D. Patel Aalpen A. Arbabshirani, Mohammad R. and Gregory J. Moore. Advanced machine learning in action: Identification of intracranial hemorrhage on computed tomography scans of the head with clinical workflow integration. *NPJ Digital Medicine*, In Press, 2018.

[8] Mohammad R Arbabshirani, Ahmed H Dallal, Chirag Agarwal, Aalpan Patel, and Gregory Moore. Accurate segmentation of lung fields on chest radiographs using deep convolutional networks. In *SPIE Medical Imaging*, pages 1013305–1013305. International Society for Optics and Photonics, 2017.

[9] Azizi, Shekoofeh and Bayat, Sharareh and Yan, Pingkun and Tahmasebi, Amir and Nir, Guy and Kwak, Jin Tae and Xu, Sheng and Wilson, Storey and Iczkowski, Kenneth A and Lucia, M Scott and others. Detection and grading of prostate cancer using temporal enhanced ultrasound: combining deep neural networks and tissue mimicking simulations. *International journal of computer assisted radiology and surgery*, 12(8):1293–1305, 2017.

[10] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep

convolutional encoder-decoder architecture for image segmentation. *arXiv preprint arXiv:1511.00561*, 2015.

[11] Paul Beard. Biomedical photoacoustic imaging. *Interface focus*, page rsfs20110028, 2011.

[12] Yoshua Bengio, Pascal Lamblin, Dan Popovici, and Hugo Larochelle. Greedy layer-wise training of deep networks. In *Advances in neural information processing systems*, pages 153–160, 2007.

[13] James Bergstra and Yoshua Bengio. Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13(Feb):281–305, 2012.

[14] Bezdek, James C and Hall, LO and Clarke, L_P. Review of mr image segmentation techniques using pattern recognition. *Medical physics*, 20(4):1033–1048, 1993.

[15] Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.

[16] Marcus D Bloice, Christof Stocker, and Andreas Holzinger. Augmentor: an image augmentation library for machine learning. *arXiv preprint arXiv:1708.04680*, 2017.

[17] Andrew P Bradley. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern recognition*, 30(7):1145–1159, 1997.

[18] Matthew C Chen, Robyn L Ball, Lingyao Yang, Nathaniel Moradzadeh, Brian E Chapman, David B Larson, Curtis P Langlotz, Timothy J Amrhein, and Matthew P Lungren. Deep learning to classify radiology free-text reports. *Radiology*, 286(3):845–852, 2017.

[19] Bhargava Chinni, Vikram Dogra, Zichao Han, Navalgund Rao, Pedro Vallejo, Wayne Knox, Julie Bentley, and Ronald Wood. Fabrication of a novel c-scan photoacoustic imaging camera. In *2015 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pages 1–4. IEEE, 2015.

[20] Bhargava Chinni, Zichao Han, Nicholas Brown, Pedro Vallejo, Tess Jacobs, Wayne Knox, Vikram Dogra, and Navalgund Rao. Multi-acoustic lens design methodology for a low cost c-scan photoacoustic imaging camera. In *Photons Plus Ultrasound: Imaging and Sensing 2016*, volume 9708, page 97081Q. International Society for Optics and Photonics, 2016.

[21] Benjamin T Cox, Simon R Arridge, and Paul C Beard. Estimating chromophore distributions from multiwavelength photoacoustic images. *JOSA A*, 26(2):443–455, 2009.

[22] Benjamin T Cox, Simon R Arridge, Kornel P Köstli, and Paul C Beard. Two-dimensional quantitative photoacoustic image reconstruction of absorption distributions in scattering media by use of a simple iterative method. *Applied Optics*, 45(8):1866–1875, 2006.

[23] Benjamin T Cox, Jan G Laufer, Paul C Beard, and Simon R Arridge. Quan-

titative spectroscopic photoacoustic imaging: a review. *Journal of biomedical optics*, 17(6):061202, 2012.

[24] Shusil Dangi, Ziv Yaniv, and Cristian Linte. A Distance Map Regularized CNN for Cardiac Cine MR Image Segmentation. *arXiv preprint arXiv:1901.01238*, 2019.

[25] Amy Berrington De Gonzalez. Estimates of the potential risk of radiation-related cancer from screening in the uk, 2011.

[26] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[27] Vikram Dogra, Saugata Sinha, Bhargava Chinni, Edward Messing, Ahmed Ghazi, and Navalgund Rao. Differentiation between malignant, benign prostatic hyperplasia and normal prostate tissue using frequency analysis of multispectral photoacoustic images. *The Journal of Urology*, 191(4), 2014.

[28] Vikram S Dogra, Bhargava K Chinni, Keerthi S Valluru, Jean V Joseph, Ahmed Ghazi, Jorge L Yao, Katie Evans, Edward M Messing, and Navalgund A Rao. Multispectral photoacoustic imaging of prostate cancer: preliminary ex-vivo results. *Journal of clinical imaging science*, 3, 2013.

[29] Vikram S Dogra, Bhargava K Chinni, Keerthi S Valluru, Jacob Moalem, Ellen J Giampoli, Katie Evans, and Navalgund A Rao. Preliminary results

of ex vivo multispectral photoacoustic imaging in the management of thyroid cancer. *American Journal of Roentgenology*, 202(6):W552–W558, 2014.

[30] V Dograa, S Bhatta, K Vallurua, B Chinnia, and N Raob. Photoacoustic spectroscopy for prostate cancer detection.

[31] Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639):115, 2017.

[32] Kalloor Joseph Francis, Bhargava Chinni, Sumohana S Channappayya, Rajalakshmi Pachamuthu, Vikram S Dogra, and Navalgund Rao. Characterization of lens based photoacoustic imaging system. *Photoacoustics*, 8:37–47, 2017.

[33] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning*, volume 1. Springer series in statistics New York, NY, USA:, 2001.

[34] Zichao Han. *The Design and Realization of a Dual Mode Photoacoustic and Ultrasound Imaging Camera*. PhD thesis, 2018.

[35] Zichao Han, Bhargava Chinni, Vikram Dogra, and Nalvagund Rao. An acoustic lens based ultrasound imaging camera. In *2016 IEEE Western New York Image and Signal Processing Workshop (WNYISPW)*, pages 1–4. IEEE, 2016.

[36] James A Hanley and Barbara J McNeil. The meaning and use of the area

under a receiver operating characteristic (roc) curve. *Radiology*, 143(1):29–36, 1982.

[37] Andreas Hauptmann, Felix Lucka, Marta Betcke, Nam Huynh, Jonas Adler, Ben Cox, Paul Beard, Sebastien Ourselin, and Simon Arridge. Model-based learning for accelerated, limited-view 3-d photoacoustic tomography. *IEEE transactions on medical imaging*, 37(6):1382–1393, 2018.

[38] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015.

[39] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[40] CGA Hoelen, FFM De Mul, R Pongers, and A Dekker. Three-dimensional photoacoustic imaging of blood vessels in tissue. *Optics letters*, 23(8):648–650, 1998.

[41] Christoph GA Hoelen and Frits FM de Mul. Image reconstruction for photoacoustic scanning of tissue structures. *Applied Optics*, 39(31):5872–5883, 2000.

[42] S-I Horikawa, Takeshi Furuhashi, and Yoshiki Uchikawa. On fuzzy model-

ing using fuzzy neural networks with the back-propagation algorithm. *IEEE transactions on Neural Networks*, 3(5):801–806, 1992.

[43] AH Hou, D Swanson, and AB Barqawi. Modalities for imaging of prostate cancer. *Advances in urology*, 2009, 2009.

[44] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456, 2015.

[45] Mandakini Jain. *Design and Development of a Novel Photoacoustic Imaging System for Detection of Intracranial Hemorrhages*. PhD thesis, 2019.

[46] Andrew Janowczyk and Anant Madabhushi. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *Journal of pathology informatics*, 7, 2016.

[47] K Jnawali, B Chinni, V Dogra, and N Rao. Photoacoustic simulation study of chirp excitation response from different size absorbers. In *Medical Imaging 2017: Ultrasonic Imaging and Tomography*, volume 10139, page 101391L. International Society for Optics and Photonics, 2017.

[48] Kamal Jnawali, Mohammad R Arbabshirani, Navalgund Rao, and Alpen A Patel. Deep 3d convolution neural network for ct brain hemorrhage classification. In *Medical Imaging 2018: Computer-Aided Diagnosis*, volume 10575, page 105751C. International Society for Optics and Photonics, 2018.

[49] Kamal Jnawali, Mohammad R Arbabshirani, Alvaro E Ulloa, Navalgund Rao, and Aalpen A Patel. Automatic classification of radiological report for intracranial hemorrhage. In *2019 IEEE 13th International Conference on Semantic Computing (ICSC)*, pages 187–190. IEEE.

[50] Kamal Jnawali, Bhargava Chinni, Vikram Dogra, and Navalgund Rao. Transfer learning for automatic cancer tissue detection using multispectral photoacoustic imaging. In *Medical Imaging 2019: Computer-Aided Diagnosis*, volume 10950, page 109503W. International Society for Optics and Photonics, 2019.

[51] Kamal Jnawali, Bhargava Chinni, Vikram Dogra, Saugata Sinha, and Navalgund Rao. Deep 3d convolutional neural network for automatic cancer tissue detection using multispectral photoacoustic imaging. In *Medical Imaging 2019: Ultrasonic Imaging and Tomography*, volume 10955, page 109551D. International Society for Optics and Photonics, 2019.

[52] Kamal Jnawali, Jhon P Kerekes, and Navalgund Rao. Comparative study of spectral matched filter, constrained energy minimization, and adaptive coherence estimator for subpixel target detection based on hyperspectral imaging. In *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XXIV*, volume 10644, page 106441V. International Society for Optics and Photonics, 2018.

[53] Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Li Fei-Fei. Large-scale video classification with convolutional

neural networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 1725–1732, 2014.

[54] Baris Kayalibay, Grady Jensen, and Patrick van der Smagt. Cnn-based segmentation of medical imaging data. *arXiv preprint arXiv:1701.03056*, 2017.

[55] Deepa Sheth Maryellen L. Giger Kayla R. Mendel, Hui Li. Transfer learning with convolutional neural networks for lesion classification on clinical breast tomosynthesis, 2018.

[56] Justin Ker, Satya P Singh, Yeqi Bai, Jai Rao, Tchoyoson Lim, and Lipo Wang. Image thresholding improves 3-dimensional convolutional neural network diagnosis of different acute brain hemorrhages on computed tomography scans. *Sensors*, 19(9):2167, 2019.

[57] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[58] Ron Kohavi et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, volume 14, pages 1137–1145. Montreal, Canada, 1995.

[59] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[60] Paras Lakhani and Baskaran Sundaram. Deep learning at chest radiography:

automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology*, 284(2):574–582, 2017.

[61] Bahman Lashkari and Andreas Mandelis. Linear frequency modulation photoacoustic radar: Optimal bandwidth and signal-to-noise ratio for frequency-domain imaging of turbid media. *The Journal of the Acoustical Society of America*, 130(3):1313–1324, 2011.

[62] Changhui Li and Lihong V Wang. Photoacoustic tomography and sensing in biomedicine. *Physics in Medicine & Biology*, 54(19):R59, 2009.

[63] Hai Li, Patrick Kumavor, Umar Salman Alqasemi, and Quing Zhu. Utilizing spatial and spectral features of photoacoustic imaging for ovarian cancer detection and diagnosis. *Journal of biomedical optics*, 20(1):016002, 2015.

[64] Qing Li, Weidong Cai, Xiaogang Wang, Yun Zhou, David Dagan Feng, and Mei Chen. Medical image classification with convolutional neural network. In *2014 13th International Conference on Control Automation Robotics & Vision (ICARCV)*, pages 844–848. IEEE, 2014.

[65] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen AWM van der Laak, Bram van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88, 2017.

[66] Geert Litjens, Clara I Sánchez, Nadya Timofeeva, Meyke Hermsen, Iris Nagtegaal, Iringo Kovacs, Christina Hulsbergen-Van De Kaa, Peter Bult, Bram

Van Ginneken, and Jeroen Van Der Laak. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Scientific reports*, 6:26286, 2016.

[67] Liu, Shengfeng and Wang, Yi and Yang, Xin and Lei, Baiying and Liu, Li and Li, Shawn Xiang and Ni, Dong and Wang, Tianfu. Deep learning in medical ultrasound analysis: A review. *Engineering*, 2019.

[68] Marc Moreno Lopez and Jugal Kalita. Deep learning applied to nlp. *arXiv preprint arXiv:1703.03091*, 2017.

[69] Mallidi, Srivalleesha and Luke, Geoffrey P and Emelianov, Stanislav. Photoacoustic imaging in cancer detection, diagnosis, and treatment guidance. *Trends in biotechnology*, 29(5):213–221, 2011.

[70] David A Mankoff, Lisa K Dunnwald, Julie R Gralow, Georgiana K Ellis, Erin K Schubert, Jeffrey Tseng, Thomas J Lawton, Hannah M Linden, and Robert B Livingston. Changes in blood flow and metabolism in locally advanced breast cancer treated with neoadjuvant chemotherapy. *Journal of Nuclear Medicine*, 44(11):1806–1814, 2003.

[71] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)*, pages 565–571. IEEE, 2016.

[72] Alexander A Oraevsky, Elena V Savateeva, Sergey V Solomatin, Alexander A Karabutov, Valeri G Andreev, Zoran Gatalica, Tuenchit Khamapirad, and P Mark Henrichs. Optoacoustic imaging of blood for visualization and diagnostics of breast cancer. In *Biomedical Optoacoustics III*, volume 4618, pages 81–95. International Society for Optics and Photonics, 2002.

[73] Manojit Pramanik, Geng Ku, Changhui Li, and Lihong V Wang. Design and evaluation of a novel breast cancer detection system combining both thermoacoustic (ta) and photoacoustic (pa) tomography. *Medical physics*, 35(6Part1):2218–2223, 2008.

[74] Lutz Prechelt. Early stopping-but when? In *Neural Networks: Tricks of the trade*, pages 55–69. Springer, 1998.

[75] Arjun Raj Rajanna, Raymond Ptucha, Saugata Sinha, Bhargava Chinni, Vikram Dogra, and Navalgund A Rao. Prostate cancer detection using photoacoustic imaging and deep learning. *Electronic Imaging*, 2016(15):1–6, 2016.

[76] Navalgund Rao, Francis Kalloor, Bhargava Chinni, Zichao Han, and Vikram Dogra. Innovative approach for including dual mode ultrasound and volumetric imaging capability within a medical photoacoustic imaging camera system. In *Optical Tomography and Spectroscopy*, pages OW4D–2. Optical Society of America, 2018.

[77] Navalgund A Rao, Di Lai, Shweta Bhatt, Stephen C Arnold, Bhargava Chinni, and Vikram S Dogra. Acoustic lens characterization for ultrasound and pho-

toacoustic c-scan imaging modalities. In *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pages 2177–2180. IEEE, 2008.

[78] Daniel Razansky, Claudio Vinegoni, and Vasilis Ntziachristos. Multispectral photoacoustic imaging of fluorochromes in small animals. *Optics letters*, 32(19):2891–2893, 2007.

[79] Razzak, Muhammad Imran and Naz, Saeeda and Zaib, Ahmad. Deep learning for medical image processing: Overview, challenges and the future. In *Classification in BioApps*, pages 323–350. Springer, 2018.

[80] Nadaparambil Aravindakshan Rejesh and Manojit Pramanik. Photoacoustic and thermoacoustic signal characteristics study. In *European Conference on Biomedical Optics*, page 88000G. Optical Society of America, 2013.

[81] Nadaparambil Aravindakshan Rejesh, Harish Pullagurla, and Manojit Pramanik. Deconvolution-based deblurring of reconstructed images in photoacoustic/thermoacoustic tomography. *JOSA A*, 30(10):1994–2001, 2013.

[82] Yair Rivenson, Zoltán Göröcs, Harun Günaydin, Yibo Zhang, Hongda Wang, and Aydogan Ozcan. Deep learning microscopy. *Optica*, 4(11):1437–1443, 2017.

[83] Jessica Ruiz, Farouk Nouizi, Jaedu Cho, Jie Zheng, Yifan Li, Jeon-Hor Chen, Min-Ying Su, and Gultekin Gulsen. Breast density quantification using

structured-light-based diffuse optical tomography simulations. *Applied optics*, 56(25):7146–7157, 2017.

[84] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 618–626, 2017.

[85] Junhui Shi, Yuqi Tang, and Junjie Yao. Advances in super-resolution photoacoustic imaging. *Quantitative imaging in medicine and surgery*, 8(8):724, 2018.

[86] Hoo-Chang Shin, Holger R Roth, Mingchen Gao, Le Lu, Ziyue Xu, Isabella Nogues, Jianhua Yao, Daniel Mollura, and Ronald M Summers. Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. *IEEE transactions on medical imaging*, 35(5):1285–1298, 2016.

[87] Shinohara, Katsuto and Wheeler, Thomas M and Scardino, Peter T. The appearance of prostate cancer on transrectal ultrasonography: correlation of imaging and pathological examinations. *The Journal of urology*, 142(1):76–82, 1989.

[88] Siegel, Rebecca L and Miller, Kimberly D and Jemal, Ahmedin. Cancer statistics, 2016. *CA: a cancer journal for clinicians*, 66(1):7–30, 2016.

[89] Saugata Sinha. *Photoacoustic Image Analysis for Cancer Detection and Building a Novel Ultrasound Imaging System*. PhD thesis, 2018.

[90] Saugata Sinha, Vikram S Dogra, Bhargava K Chinni, and Navalgund A Rao. Frequency domain analysis of multiwavelength photoacoustic signals for differentiating among malignant, benign, and normal thyroids in an ex vivo study with human thyroids. *Journal of Ultrasound in Medicine*, 36(10):2047–2059, 2017.

[91] Saugata Sinha, Navalgund A Rao, Keerthi S Valluru, Bhargava K Chinni, Vikram S Dogra, and Maria Helguera. Frequency analysis of multispectral photoacoustic images for differentiating malignant region from normal region in excised human prostate. In *Medical Imaging 2014: Ultrasonic Imaging and Tomography*, volume 9040, page 90400P. International Society for Optics and Photonics, 2014.

[92] Jasper Snoek, Hugo Larochelle, and Ryan P Adams. Practical bayesian optimization of machine learning algorithms. In *Advances in neural information processing systems*, pages 2951–2959, 2012.

[93] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958, 2014.

[94] John C Stover. *Optical scattering: measurement and analysis*, volume 2. SPIE optical engineering press Bellingham, 1995.

[95] Wenqing Sun, Bin Zheng, and Wei Qian. Computer aided lung cancer diagnosis with deep learning algorithms. In *Medical imaging 2016: computer-aided diagnosis*, volume 9785, page 97850Z. International Society for Optics and Photonics, 2016.

[96] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *AAAI*, volume 4, page 12, 2017.

[97] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[98] Nima Tajbakhsh, Jae Y Shin, Suryakanth R Gurudu, R Todd Hurst, Christopher B Kendall, Michael B Gotway, and Jianming Liang. Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE transactions on medical imaging*, 35(5):1299–1312, 2016.

[99] Adrian Taruttis and Vasilis Ntziachristos. Advances in real-time multispectral optoacoustic imaging and its applications. *Nature photonics*, 9(4):219, 2015.

[100] Bradley E Treeby and Benjamin T Cox. k-wave: Matlab toolbox for the sim-

ulation and reconstruction of photoacoustic wave fields. *Journal of biomedical optics*, 15(2):021314, 2010.

[101] Tsehay, Yohannes K and Lay, Nathan S and Roth, Holger R and Wang, Xiaosong and Kwak, Jin Tae and Turkbey, Baris I and Pinto, Peter A and Wood, Brad J and Summers, Ronald M. Convolutional neural network based deep-learning architecture for prostate cancer detection on multiparametric magnetic resonance images. In *Medical Imaging 2017: Computer-Aided Diagnosis*, volume 10134, page 1013405. International Society for Optics and Photonics, 2017.

[102] Paul Kumar Upputuri and Manojit Pramanik. Performance characterization of low-cost, high-speed, portable pulsed laser diode photoacoustic tomography (pld-pat) system. *Biomedical optics express*, 6(10):4118–4129, 2015.

[103] Keerthi S Valluru, Bhargava K Chinni, and Navalgund A Rao. Photoacoustic imaging: opening new frontiers in medical imaging. *Journal of clinical imaging science*, 1, 2011.

[104] Keerthi S Valluru, Bhargava K Chinni, Navalgund A Rao, Shweta Bhatt, and Vikram S Dogra. Basics and clinical applications of photoacoustic imaging. *Ultrasound Clinics*, 4(3):403–429, 2009.

[105] Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan. Show and tell: A neural image caption generator. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3156–3164, 2015.

[106] Alfred Vogel and Vasan Venugopalan. Mechanisms of pulsed laser ablation of biological tissues. *Chemical reviews*, 103(2):577–644, 2003.

[107] Lihong V Wang. Ultrasound-mediated biophotonic imaging: a review of acousto-optical tomography and photo-acoustic tomography. *Disease markers*, 19(2, 3):123–138, 2004.

[108] Lihong V Wang. Tutorial on photoacoustic microscopy and computed tomography. *IEEE Journal of Selected Topics in Quantum Electronics*, 14(1):171–179, 2008.

[109] Lihong V Wang. Photoacoustic tomography: High-resolution imaging of optical contrast in vivo at superdepths. In *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1201–1201. IEEE, 2009.

[110] Lihong V Wang and Song Hu. Photoacoustic tomography: in vivo imaging from organelles to organs. *science*, 335(6075):1458–1462, 2012.

[111] Xueding Wang, Yongjiang Pang, Geng Ku, George Stoica, and Lihong V Wang. Three-dimensional laser-induced photoacoustic tomography of mouse brain with the skin and skull intact. *Optics letters*, 28(19):1739–1741, 2003.

[112] Xueding Wang, Yongjiang Pang, Geng Ku, Xueyi Xie, George Stoica, and Lihong V Wang. Noninvasive laser-induced photoacoustic tomography for structural and functional in vivo imaging of the brain. *Nature biotechnology*, 21(7):803, 2003.

[113] Yi Wang, Da Xing, Yaguang Zeng, and Qun Chen. Photoacoustic imaging with deconvolution algorithm. *Physics in Medicine & Biology*, 49(14):3117, 2004.

[114] Jiajun Wu, Yinan Yu, Chang Huang, and Kai Yu. Deep multiple instance learning for image classification and auto-annotation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3460–3469, 2015.

[115] Jun Xia, Junjie Yao, and Lihong V Wang. Photoacoustic tomography: principles and advances. *Electromagnetic waves (Cambridge, Mass.)*, 147:1, 2014.

[116] Minghua Xu and Lihong V Wang. Photoacoustic imaging in biomedicine. *Review of scientific instruments*, 77(4):041101, 2006.

[117] Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *European conference on computer vision*, pages 818–833. Springer, 2014.

[118] Zhi-Hua Zhou and Xu-Ying Liu. Training cost-sensitive neural networks with methods addressing the class imbalance problem. *IEEE Transactions on Knowledge and Data Engineering*, 18(1):63–77, 2006.