

DEEP LEARNING MODELS FOR DETECTING GALLBLADDER CANCER FROM ULTRASOUND

SOUMEN BASU



**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY DELHI
JULY 2024**

DEEP LEARNING MODELS FOR DETECTING GALLBLADDER CANCER FROM ULTRASOUND

by

SOUMEN BASU

Department of Computer Science and Engineering

Submitted

in fulfillment of the requirements of the degree of
Doctor of Philosophy

to the



Indian Institute of Technology Delhi
JULY 2024

Certificate

This is to certify that the thesis titled “**Deep Learning Models for Detecting Gallbladder Cancer from Ultrasound**” being submitted by **Mr. Soumen Basu** for the award of **Doctor of Philosophy in Computer Science and Engineering** is a record of bona fide work carried out by him under my guidance and supervision at the Department of Computer Science and Engineering, Indian Institute of Technology Delhi. The work presented in this thesis has not been submitted elsewhere, either in part or full, for the award of any other degree or diploma.

Prof. Chetan Arora

Professor

Department of Computer Science and Engineering

Indian Institute of Technology Delhi

New Delhi, 110016

Acknowledgements

First and foremost, I extend my profound gratitude to my advisor, Professor Chetan Arora, who has not only been a brilliant mentor but also the source of motivation throughout my journey. His unwavering support has always been a constant by my side, guiding me through both academic challenges and life's uncertainties. I deeply value his brilliant research advice and insightful discussions, without which this work would not have been achievable. I owe a debt of gratitude to him for believing in me during the initial tumultuous days of my PhD.

I am indebted to Dr. Pankaj Gupta from PGIMER, Chandigarh, who has been a truly motivating clinician-cum-academic. His consistent efforts, collaboration, and enriching discussions on clinical aspects have played a pivotal role in shaping and realizing the work. I am also thankful to my SRC members, Prof. Prem Kalra, Prof. Parag Singla, and Prof. Rohan Paul for their insightful suggestions regarding improving the work. I am also thankful to CSE staff – Ms. Rekha, Mr. Hemant, Mr. Vikash and the Department of CSE for their continual support throughout my journey.

I extend heartfelt thanks to the bright and talented co-authors – Mayank Gupta, Dr. Pratyaksha Rana, Somanshu Singla, Mayuna Gupta, Chetan Madan, and Ashish Papanai – whose contributions have been instrumental. I also thank Nikhil, Ankita, Ramya, Anupam, Britty, Kunal, Ajay, Rohan, Prashant, and other vision group students for the enriching conversations from time to time. Additionally, I thank my friends, Vipul, Aritra, Argha, Arpan, Arkajyoti, Shivani, Sahil, Swaraj, Soham, Suvo, Avik, Arijit, Shamik, Biman, and many others for their support and cheerful company during the exhausting phases of this journey.

I want to express my deepest gratitude to all of my teachers who have played a pivotal role in

shaping my journey starting from kindergarten to school, through universities. Thank you for inspiring and supporting me, and making me who I am today.

I am profoundly grateful to my family, especially my wife, Arunima, my son, Syamantak, and my father, Mr. Jyotirmoy Basu, who have been the consistent pillars of support throughout this journey. They have been relentlessly there for me through thick and thin, and no amount of gratitude is enough. I am equally grateful to my mother-in-law, Mrs. Suparna Maitra, for her constant motivation.

I am forever indebted to two dear ones watching me from their heavenly abode: my mother, Mrs. Leena Basu, and my father-in-law, Dr. Ashok Maitra. Without their dream, motivation, and support, I would not have stepped into the journey of a PhD. Finally, I humbly bow down to the supreme lord for showering all the love and blessings throughout all this time.

Soumen Basu

Abstract

Gallbladder Cancer (GBC) is the most common biliary tract cancer and the 5th most common gastrointestinal tract malignancy. India sees about 20% of annual GBC-related deaths worldwide and faces an incidence rate compared to the global highest. The overall mean survival rate for patients with advanced GBC is only six months, and the 5-year survival rate is less than 5%. Early diagnosis and curative surgical resection remain the only hope to improve the bleak survival statistics. Ultrasound (USG) is a popular and excellent candidate diagnostic modality for abdominal ailments in low-resource settings due to its low cost, availability, and ionizing radiation-free nature. USG is also the first-line diagnostic modality for gallbladder (GB) diseases. However, diagnosing GBC in USG is difficult, even for experienced radiologists, due to the overlapping visual features of benign and malignant GBs and various confounding medical conditions such as cholecystitis, pancreatitis, and Rokitansky-Aschoff sinuses. Our experiments reveal that human experts could achieve only about 70% sensitivity (recall) in differentiating GBC from benign diseases (dichotomous classification) from USG.

Inspired by the recent success and the transformational capabilities shown by Machine Learning (ML) models, especially the Deep Neural Networks (DNNs), in a plethora of medical image computing tasks, we investigate leveraging DNNs to detect GBC from USG. However, the low image quality arising from noise and artifacts such as shadows or textures, the operator bias and variation in view due to handheld sensors, and the lack of annotated data make the application of DNNs difficult in USG.

In our pursuit of enhancing diagnostic capabilities, we systematically explore the potential of deep convolutional models, leading to the development of GBCNet. GBCNet is a two-stage

DNN that first localizes the GB or the region-of-interest (ROI), and then employs a specialized classifier based on multi-scale, second-order pooling (MS-SoP) for robust GBC detection. We further develop a Gaussian smoothing-based training curriculum inspired by human visual acuity to mitigate the effect of spurious textures. GBCNet tackles issues such as noise, artifacts, and viewpoint variation in USG imaging, improving the GBC detection sensitivity by 7 points compared to SOTA DNN models and 20 points compared to expert radiologists.

The reliance on bounding box annotations for training GBCNet’s localization component presents a significant bottleneck, given the high cost and complexity of obtaining such annotations. To overcome this challenge, we utilize limited supervised data by introducing – (1) an unsupervised contrastive framework for learning GBC representations from unlabelled videos, and (2) a weakly supervised object detection technique to use only image labels instead of bounding box annotation requirements, thus designing more practical models for real-world deployment.

We address the crucial aspect of interpretability in GBC detection by introducing RadFormer, a deep neural network architecture capable of generating interpretable explanations for its decisions. RadFormer not only improves detection sensitivity over GBCNet but also aids in understanding the underlying visual features relevant to GBC diagnosis, bridging the gap between AI-based detection and clinical interpretability.

Finally, we advocate for a paradigm shift towards video-based GBC detection, leveraging the rich spatiotemporal information available in full USG videos. Video-based detection is also clinically more relevant as single frames may not contain conclusive evidence for disease detection. We introduce an innovative masked auto-encoder design called FocusMAE, to learn self-supervised representations for GBC from USG videos. We demonstrate significant improvements in GBC detection using FocusMAE, achieving a 100% sensitivity and thus showcasing the potential of video-based approaches in streamlining the detection process and reducing operator-specific variations.

In summary, we designed and developed accurate, data-efficient, interpretable, and clinically relevant DNN models which could overcome the challenges such as noise, artifacts, viewpoint variability, data scarcity, and real-time applicability in detecting GBC from USG, thereby opening future avenues for transformational research.

सार

गॉलब्लैडर कैंसर (जीबीसी) सबसे सामान्य बाइलियरी ट्रैक्ट कैंसर और पांचवीं सबसे सामान्य गैस्ट्रोइंटेस्टाइनल ट्रैक्ट मेलिगनेंसी है। इंडिया में विश्वव्यापी जीबीसी-संबंधित वार्षिक मौतों का लगभग २०% हिस्सा है और यह वैश्विक उच्चतम की तुलना में उच्चतम घटना दर का सामना करता है। उन्नत जीबीसी वाले पेशेंट्स के लिए कुल औसत सर्वाइवल रेट केवल छह महीने हैं, और ५-वर्ष सर्वाइवल रेट ५% से कम है। अल्ट्रासाउंड (यूएसजी) अपनी कम लागत, उपलब्धता और आयोनाइजिंग रेडिएशन-फ्री नेचर के कारण कम रिसोर्स सेटिंग्स में एब्डॉमिनल एलमेंट्स के लिए एक पॉपुलर और एक्सीलेंट कैंडिडेट डायग्नोस्टिक मोडालिटी है। यूएसजी भी गॉलब्लैडर (जीबी) डिसीजेस के लिए पहली लाइन डायग्नोस्टिक मोडालिटी है। हालांकि, यूएसजी में जीबीसी का डायग्नोसिस करना कठिन है, यहां तक कि एक्सपीरियेंस्ड रेडियोलॉजिस्ट्स के लिए भी, क्योंकि बिनाइन और मैलिगनेंट जीबी की ओवरलैपिंग विजुअल फीचर्स और कोलेसिस्टाइटिस, पैक्रिएटाइटिस, और रोकीटांस्की-आशोफ साइनसेस जैसी विभिन्न कंपाउंडिंग मेडिकल कंडिशन के बीच ओवरलैपिंग होती हैं। हमारे एक्सपेरिमेंट्स से पता चला है कि ह्यूमन एक्सपर्ट्स यूएसजी से बिनाइन डिसीजेस से जीबीसी का अंतर करने में केवल लगभग ७०% सेंसिटिविटी (रिकॉल) ही प्राप्त कर सकते हैं।

मशीन लर्निंग (एमएल) मॉडल्स, विशेष रूप से डीप न्यूरल नेटवर्क्स (डीएनएन), द्वारा हाल के सफलता और परिवर्तनकारी क्षमताओं से प्रेरित होकर, हम यूएसजी से जीबीसी का पता लगाने के लिए डीएनएन का लाभ उठाने की जांच करते हैं। हालांकि, नॉइज़ और आर्टिफैक्ट्स जैसे शैडोज़ या टेक्चर्स से उत्पन्न निम्न इमेज क्वालिटी, हैंडहेल्ड सेंसर्स के कारण ऑपरेटर बायस और व्यू में वेरिएशन, और एनोटेटेड डेटा की कमी डीएनएन का उपयोग यूएसजी में कठिन बना देती है।

हमारे डायग्नोस्टिक क्षमताओं को बढ़ाने के प्रयास में, हम गहन कॉन्वोल्यूशनल मॉडल्स की संभावनाओं का व्यवस्थित रूप से अन्वेषण करते हैं, जिसके परिणामस्वरूप जीबीसीनेट का विकास होता है। जीबीसीनेट एक दो-स्टेज डीएनएन है जो पहले जीबी या रीजन-ऑफ-इंटरेस्ट (आरओआई) का लोकलाइजेशन करता है, और फिर रॉबस्ट जीबीसी डिटेक्शन के लिए मल्टी-स्केल, सेकंड-ऑर्डर पूलिंग (एमएस-एसओपी) पर आधारित एक विशेष क्लासिफायर का उपयोग करता है। हम स्प्यूरियस टेक्चर्स के प्रभाव को कम करने के लिए ह्यूमन विजुअल एक्क्यूइटी से प्रेरित गॉसियन स्मूदिंग-आधारित ट्रेनिंग करिकुलम भी डेवलप करते हैं। जीबीसीनेट यूएसजी इमेजिंग में नॉइज़, आर्टिफैक्ट्स, और व्यूपॉइंट वेरिएशन जैसी समस्याओं को हल करता है, जीबीसी डिटेक्शन की सेंसिटिविटी को सोता डीएनएन मॉडल्स की तुलना में ७ पॉइंट्स और एक्सपर्ट रेडियोलॉजिस्ट्स की तुलना में २० पॉइंट्स बढ़ाता है।

जीबीसीनेट के लोकलाइजेशन कंपोनेंट के ट्रेनिंग के लिए बाउंडिंग बॉक्स अनोटेशन पर निर्भरता एक महत्वपूर्ण बाधा प्रस्तुत करती है, क्योंकि इस तरह के अनोटेशन प्राप्त करना उच्च लागत और जटिलता से भरा होता है। इस चुनौती को दूर करने के लिए, हम सीमित सुपरवाइज्ड डेटा का उपयोग करने में गहराई से जुड़ते हैं और – (१) अनलेबल्ड वीडियो डेटा से जीबीसी रेप्रेजेंटेशन सीखने के लिए एक अनसुपरवाइज्ड कॉन्ट्रास्टिव फ्रेमवर्क, और (२) बाउंडिंग बॉक्स अनोटेशन की आवश्यकताओं के बजाय केवल इमेज-लेवल लेबल्स का उपयोग करने के लिए वीकली सुपरवाइज्ड ऑब्जेक्ट डिटेक्शन टेक्नीक प्रस्तुत करते हैं, इस प्रकार रियल वर्ल्ड डिप्लॉयमेंट के लिए अधिक प्रैक्टिकल मॉडल्स डिजाइन करते हैं।

हम जीबीसी डिटेक्शन में इंटरप्रेटेबिलिटी के महत्वपूर्ण पहलू को रैंडफॉर्म के परिचय के साथ संबोधित करते हैं, जो एक गहन तंत्रिका नेटवर्क आर्किटेक्चर है जो अपने निर्णयों के लिए इंटरप्रेटेबिल स्पष्टीकरण उत्पन्न करने में सक्षम है। रैंडफॉर्म न केवल जीबीसीनेट पर डिटेक्शन सेंसिटिविटी में सुधार करता है बल्कि जीबीसी निदान के लिए प्रासंगिक अंतर्निहित विजुअल फीचर्स को समझने में भी मदद करता है, एआई-आधारित डिटेक्शन और क्लिनिकल इंटरप्रेटेबिलिटी के बीच की खाई को पाटता है।

अंत में, हम पूर्ण यूएसजी वीडियो में उपलब्ध रिच स्पैटिओटेम्पोरल इंफॉर्मेशन का लाभ उठाते हुए वीडियो-बेस्ड जीबीसी डिटेक्शन की ओर एक पैराडाइम शिफ्ट की वकालत करते हैं। वीडियो-बेस्ड डिटेक्शन नैदानिक रूप से अधिक प्रासंगिक है क्योंकि सिंगल फ्रेम में डिजीज का पता लगाने के लिए निर्णायक सबूत नहीं हो सकते हैं। हम जीबीसी के लिए यूएसजी वीडियो से सेल्फ-सुपरवाइज्ड रेप्रेजेंटेशन सीखने के लिए एक इनोवेटिव मास्कड ऑटो-एन्कोडर डिजाइन जिसे फोकसएमआई कहते हैं, पेश करते हैं। हम फोकसएमआई का उपयोग करके जीबीसी डिटेक्शन में महत्वपूर्ण सुधार दिखाते हैं, १००% सेंसिटिविटी प्राप्त करते हैं और इस प्रकार वीडियो-बेस्ड अप्रोचेस की संभावनाओं को प्रदर्शित करते हैं जो डिटेक्शन प्रक्रिया को सुव्यवस्थित करते हैं और ऑपरेटर-स्पेसिफिक वेरिएशन्स को कम करते हैं।

सारांश में, हमने सटीक, डेटा-कुशल, इंटरप्रेटेबल, और क्लिनिकली प्रासंगिक डीएनएन मॉडल्स डिजाइन और डेवलप किए हैं जो यूएसजी से जीबीसी का पता लगाने में नॉइज़, आर्टिफैक्ट्स, व्यूपॉइंट वेरिएबिलिटी, डेटा स्कार्सिटी, और रियल-टाइम अप्लिकेबिलिटी जैसी चुनौतियों को पार कर सकते हैं, इस प्रकार भविष्य के परिवर्तनकारी अनुसंधान के लिए नए रास्ते खोलते हैं।

Contents

Certificate	i
Acknowledgements	iii
Abstract	v
List of Figures	xiii
List of Tables	xvii
1 Introduction	1
1.1 Background	1
1.2 Gallbladder Cancer and its Risks	3
1.2.1 Gallbladder Cancer	3
1.2.2 Risks of GBC	4
1.3 Motivation of the Study	4
1.4 Challenges of using Ultrasound	5
1.5 Key Research Objectives	6
1.6 Organization of Subsequent Chapters	7
2 Ultrasound and its Challenges in GBC Detection	9
2.1 Basics of Ultrasound	9

2.1.1	Ultrasound Physics	9
2.1.2	Common Modes of USG	10
2.1.3	Types of Probes	12
2.1.4	Body Planes	13
2.2	Brief Description of Abdominal USG	14
2.3	Brief Specification of Our Data	15
2.4	Noise and Artifacts in Abdominal Ultrasound	15
2.5	Major Challenges	17
2.5.1	Challenges in Clinical Characterization of GB Wall Thickening	17
2.5.2	Low Image Quality – Noise and Artifacts	18
2.5.3	Handheld Transducer – Viewpoint and Operator Variability	19
2.5.4	Low Inter-Class and High Intra-Class Variability	21
2.5.5	Non-regular Anatomy	22
2.5.6	Effect of the Metabolic State of the Patients	23
2.5.7	Lack of Annotated Data	24
2.5.8	Challenges Addressed in this Work	24
3	Datasets	27
3.1	Data Collection	27
3.2	Image-based Dataset	28
3.2.1	Labeling and ROI Annotation	28
3.2.2	Dataset Statistics	29
3.2.3	Dataset Splits	29
3.3	Video Data	30
3.3.1	Collection and Curation	30
3.3.2	Dataset Statistics	31
3.3.3	Dataset Splits for Video-based GBC Detection	31

4	Developing an Accurate GBC Detector	33
4.1	Introduction	34
4.2	Related Work	37
4.3	Our Method	39
4.3.1	GBCNet: Model Architecture	39
4.3.2	Visual Acuity Inspired Curriculum	42
4.4	Implementation Details	45
4.4.1	Transfer Learning and Data Augmentation	45
4.4.2	Hyper-parameters	46
4.5	Evaluation Metrics	46
4.6	Experiments and Results	47
4.6.1	Dataset	47
4.6.2	Efficacy of GBCNet over Baselines	49
4.6.3	Performance of GB Region Selection Models	51
4.6.4	Applicability of the Proposed Classifier in Breast Cancer Detection from USG Images	52
4.6.5	Efficacy of the Proposed Curriculum	52
4.6.6	Ablation Study	54
4.7	Conclusion	56
5	Learning from Limited Supervised Data I: Using Unlabelled Videos	57
5.1	Introduction	58
5.2	Data	60
5.2.1	USG Dataset for Gallbladder Cancer	60
5.2.2	Public Lung USG Dataset for COVID-19	61
5.3	Methodology	61
5.4	Experiments and Results	64
5.4.1	Experimental Setup	64

5.4.2	Comparison with Baselines	65
5.4.3	Generality of our Method Across Datasets	67
5.4.4	Ablation Study	67
5.5	Conclusion	69
6	Learning from Limited Supervised Data II: Weakly Supervised GBC Detection	71
6.1	Introduction	72
6.2	Data	74
6.2.1	GBC Detection in USG Images	74
6.2.2	Polyp Detection in Colonoscopy Images	74
6.3	Our Method	74
6.4	Experiments and Results	77
6.4.1	Experimental Setup	77
6.4.2	Comparison with Baseline WSOD Methods	77
6.4.3	Generality of the Method	78
6.4.4	Ablation Study	79
6.4.5	Classification Performance	79
6.5	Conclusion	80
7	Interpretable GBC Detection	81
7.1	Introduction	82
7.2	Related Work	83
7.3	Dataset	86
7.4	Methodology	87
7.4.1	Model Architecture	87
7.4.2	Network training	93
7.4.3	Explainability of the network inferences	94
7.5	Experiments and results	97

7.5.1	Performance of RadFormer in predicting GBC	97
7.5.2	Interpretability of predictions	100
7.5.3	Discovery of unknown malignancy indicators	100
7.5.4	Ablation study	102
7.6	Discussion and conclusion	105
8	Video-based Detection of GBC	111
8.1	Introduction	112
8.2	Related Work	115
8.3	Proposed Method	116
8.3.1	Object Priors in MAE	116
8.3.2	FocusMAE Architecture	117
8.3.3	Training	120
8.4	Dataset	121
8.4.1	Curated USG Video Dataset for GBC Detection	121
8.4.2	Public CT Dataset for Covid Detection	122
8.5	Implementation and Evaluation	122
8.6	Experiments and Results	124
8.6.1	Efficacy of FocusMAE over SOTA Baselines	124
8.6.2	Generality of the Proposed Method	127
8.6.3	Ablation Study	128
8.6.4	Analysis on Candidate Region Selection	130
8.7	Conclusion	130
9	Conclusion and Future Directions	133
9.1	Limitations and Scope of the Work	135
9.1.1	Limitations	135
9.1.2	Scope of the Work and Generalizability	135

9.2	Future Directions	136
9.2.1	Translation to Clinical Settings	136
9.2.2	Domain Adaptation/ Generalization – Multi-Center Study	137
9.2.3	Applicability of Foundation Models	138
9.2.4	Self-supervised Learning for USG	138
9.2.5	Multimodal Learning	138
9.2.6	Discovering Hidden Features with AI	139
9.3	Closing Remarks	139
 Bibliography		 141
 List of Publications		 165
 Biography		 167

List of Figures

1.1	Anatomy of Gallbladder in the human body	2
1.2	The worldwide Age Standardized Incidence Rates (ASR) of GBC	3
1.3	Ultrasound image samples	6
2.1	Sample USG images from different modes	11
2.2	Common USG probes	12
2.3	Different body planes for radiological scanning	13
2.4	Ultrasound machine, probe, and schematic diagram of abdominal USG	14
2.5	Visualization of common artifacts present in our data	16
2.6	Appearances of benign and malignant wall thickening	18
2.7	Visualization of issues arising from artifacts such as acoustic shadow	19
2.8	Textures of liver tissues leading to incorrect GBC predictions	20
2.9	Visualization of DNN models getting biased by hyperechoic regions	20
2.10	Angle of transducer causing pseudo structures	21
2.11	Viewpoint variation leading to incorrect diagnosis	21
2.12	Low inter-class and high intra-class variability of GBC in USG data	22
2.13	Visualization of non-regular anatomy of the malignant gallbladders	22
2.14	Effect of metabolic state on GB distension	23
3.1	Sample ROI annotation in USG images	29
3.2	Video data samples	30

4.1	Depiction of issues with USG-based GBC detection	36
4.2	Architecture of GBCNet	39
4.3	Visualization of texture bias	43
4.4	Gaussian blurring of USG data samples	43
4.5	Qualitative analysis of GBCNet with Grad-CAM visuals	48
4.6	Additional Grad-CAM visuals for GBCNet	49
4.7	Comparison between the ROIs predicted by Faster-RCNN with radiologists . .	50
4.8	Additional ROI visuals	51
4.9	Visualization of the effect of curriculum	54
4.10	Ablation study	55
5.1	Motivation of using intra-video negatives	58
5.2	Sample video sequences for the GBC and Covid Ultrasound data	60
5.3	Overview of the proposed contrastive framework	62
5.4	Grad-CAM visuals of the models with ImageNet pretraining and our contrastive pretraining	66
5.5	Sensitivity of pretraining hyper-parameters	69
6.1	Visuals of GBC and Polyp data samples	73
6.2	Overview of the proposed Weakly Supervised DETR architecture	75
6.3	Qualitative analysis of the predicted bounding boxes	79
7.1	Schematic diagram of the RadFormer architecture	87
7.2	Generation of the ROI and the local patches	88
7.3	Human-readable lexicons identified by RadFormer	95
7.4	ROC curves	99
7.5	Sample explanation of the network decisions	101
7.6	Visualization of unknown feature	103
7.7	Ablation on the transformer layers	105

7.8	Visualization of a malignant feature	106
7.9	Visualization of a benign feature	107
7.10	Frequency of top-10 activated features	108
7.11	Visuals of few top-activated unmatched features	109
8.1	Masking strategy of FocusMAE and its motivation	113
8.2	Overview of the FocusMAE pipeline	117
8.3	Sample video sequences from the USG and CT datasets	121
8.4	Visual demonstration of the benefit of using the FocusMAE	125
8.5	Attention visuals for FocusMAE for GBC detection	126
8.6	Attention visuals for FocusMAE for Covid detection	127
8.7	Ablation study on FocusMAE	128
8.8	Visuals of candidate region priors	130

List of Tables

4.1	Implementation details for the different baseline networks	45
4.2	Comparing GBCNet with baselines for detecting GBC from USG images . . .	47
4.3	Comparison of the GB region selection models	50
4.4	Breast cancer detection results	52
4.5	Robustness of the visual acuity curriculum in tackling texture bias	53
4.6	Model performances for training with visual acuity curriculum	54
4.7	Ablation of the proposed MS-SoP network	56
5.1	Comparison of the proposed contrastive framework with SOTA on GBC detection	64
5.2	Performance of the proposed method in Covid detection	64
5.3	Comparison with the radiologists	65
5.4	Quantitative validation of joint mining of intra, and cross video negative	65
5.5	Effectiveness of the curriculum-based negative mining	67
5.6	Performance of different backbones with our contrastive framework	68
6.1	Weakly supervised disease detection performance of SOTA and our method . .	78
6.2	Ablation study on the proposed weakly supervised DETR	78
6.3	comparison of the proposed WSOD and other SOTA methods in GBC classification	80
7.1	Performance comparison of RadFormer with SOTA	98

7.2	Comparison of RadFormer with the radiologists	100
7.3	Ablation on binarization	104
7.4	Ablation on design components	104
7.5	Choice of backbone in global branch	105
8.1	Comparison of SOTA and FocusMAE for GBC detection	124
8.2	Comparison of SOTA and FocusMAE for Covid detection	127
8.3	Hyperparameter tuning of FocusMAE	129

