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 मेलिग्नेंसी 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 है, और ५-वर्ष सर्वाइवल रेट ५% से कम है। अर्ली डायग्नोसिस और क्युरेटिव सर्जिकल रेजेक्शन ही दुखद सर्वाइवल स्टेटिस्टिक्स को सुधारने की एकमात्र आशा है। अल्ट्रासाउंड (यूएसजी) अपनी कम लागत, उपलब्धता और आयोनाइज़िंग रेडिएशन-फ्री नेचर के कारण कम रिसोर्स सेटिंग्स में एब्डॉमिनल एल्मेंट्स के लिए एक पॉपुलर और एक्सीलेंट कैंडिडेट डायग्नोस्टिक मोडालिटी है। यूएसजी भी गॉलब्लैडर (जीबी) डिसीज़ेस के लिए पहली लाइन डायग्नोस्टिक मोडालिटी है। हालांकि, यूएसजी में जीबीसी का डायग्नोसिस करना कठिन है, यहां तक कि एक्सपीरियंस्ड रेडियोलॉजिस्ट्स के लिए भी, क्योंकि बिनाइन और मैलिग्नेंट जीबी की ओवरलैपिंग विजुअल फीचर्स और कोलेसिस्टाइटिस, पैंक्रिएटाइटिस, और रोकीटांस्की-आशोफ साइनसेस जैसी विभिन्न कंफाउंडिंग मेडिकल कंडिशन्स के बीच ओवरलैपिंग होती हैं। हमारे एक्सपेरिमेंट्स से पता चला है कि ह्यूमन एक्सपर्ट्स यूएसजी से बिनाइन डिसीज़ेस 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 टेक्स्चर्स 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 कॉनवोल्यूशनल 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 टेक्स्चर्स के प्रभाव को कम करने के लिए ह्यूमन विजुअल एक्यूइटी 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 अनोटेशन्स पर निर्भरता एक महत्वपूर्ण बाधा प्रस्तुत करती है, क्योंकि इस तरह के अनोटेशन्स 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 रेप्रेजेंटेशन्स सीखने के लिए एक अनसुपरवाइज्ड कॉन्ट्रास्टिव फ्रेमवर्क, और (२) बाउंडिंग बॉक्स अनोटेशन्स 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 रैडफॉर्मर 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। रैडफॉर्मर 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 रेप्रेजेंटेशन्स सीखने के लिए एक इनोवेटिव मास्क्ड ऑटो-एन्कोडर डिजाइन जिसे फोकसएमएई कहते हैं, पेश करते हैं। हम फोकसएमएई 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

List of Figures

List of Tables