UNIVERSITY
OF TECHNOLOGY
SYDNEY

Rapid Detection and Prognosis of Lung Cancer

by Yujiao Wu

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

under the supervision of Steven Su

University of Technology Sydney
Faculty of Engineering and Information Technology

February 2020

Certificate of Original Authorship

Required wording for the certificate of original authorship

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, *Yujiao Wu* declare that this thesis, is submitted in fulfilment of the requirements for the award of *Doctor of Philosophy*, in the *Faculty of Engineering and Information Technology* at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.
**If applicable, the above statement must be replaced with the collaborative doctoral degree statement (see below).*

**If applicable, the Indigenous Cultural and Intellectual Property (ICIP) statement must be added (see below).*

This research is supported by the Australian Government Research Training Program.

Production Note:
Signature: Signature removed prior to publication.

Date: 17/02/2021

Collaborative doctoral research degree statement

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of the requirements for a degree at any other academic institution except as fully acknowledged within the text. This thesis is the result of a Collaborative Doctoral Research Degree program with *[insert collaborative partner institution]*.

Indigenous Cultural and Intellectual Property (ICIP) statement

This thesis includes Indigenous Cultural and Intellectual Property (ICIP) belonging to *[insert relevant language, tribal or nation group(s) or communities]*, custodians or traditional owners. Where I have used ICIP, I have followed the relevant protocols and consulted with appropriate Indigenous people/communities about its inclusion in my thesis. ICIP rights are Indigenous heritage and will always remain with these groups. To use, adapt or reference the ICIP contained in this work, you will need to consult with the relevant Indigenous groups and follow cultural protocols.

Acknowledgements

Firstly, I would like to take this opportunity to express my sincere and deep gratitude to my supervisor, Associate Professor Steven Su, for his continual support, help, and encouragement during my Ph.D. study. Dr. Su has brought me into the topic of electronic nose and deep learning and provided brilliant insights into my research works. It is my honor to have a supervisor who always inspires me to achieve higher targets. His conscientious and meticulous attitude on research has a significant influence on my work.

I am grateful to my co-supervisor, Dr. Steve Ling, for his patient and constructive suggestions on my research. I feel so fortunate to have the precious chance to work with Dr. Steve. He has been significant in all aspects during my Ph.D. studies. I am deeply inspired by his hardworking and professional attitude to work. I feel really thankful for all his support and selfless help.

An extraordinary thanks go to my external supervisor Dr. Fan Yang, for his support and encouragement. His warm personality, broad vision, depth of knowledge, and dedication to work have been a true inspiration to my work and life.

I also want to express my appreciation to the staff members in the Centre for Health Technologies, School of biomedical engineering, University of Technology Sydney, especially to Professor Joanne Tipper, Professor Gyorgy Hutvagner, and Dr. Jan Szymanski. I have received a unique and memorizable experience working with these professional and inclusive people.

I would also like to thank my colleagues in A/Prof. In particular, Steven Su's research group, Dr. Lin Ye, Dr. Hairong Yu, Dr. Kairui Guo, Dr. Yao Huang, Taopin Liu, Miao Zhang, and Lingmeng Li, for their great help and technical support. I would also be grateful to my friends, in particular, Dr. Wenwei Mo, Dr. Lang Chen, Dr. Pengfei Cui, Dr. Zhichao Sheng, Dr. Haimin Zhang, Zhiyuan Shi, Xiaoshui Huang, Juan Lyu, Yimeng Feng, Hanjie Wu, and Dr.Daniel Roxby, and all of my colleagues whose name has not been mentioned here, for their valuable and worthless help. Working together with them brings a lot of happiness and a wonderful memory for me.

I want to acknowledge the financial support for this project, provided jointly by the University of Technology Sydney (UTS) and the Shenzhen ET group.

Lastly, my deepest gratitude goes to my mom, dad, uncle, grandma, and all my family, for their encouragement and support. I would also like to express my great gratitude to my grandpa; I really hope you could witness all the important moments in my life. Pale words are hard to express my gratitude to all of you. This dissertation wouldn't exist without your love.

Yujiao Wu

Sydney, Australia, Feb 2021

List of Publications

I have published one journal paper, one conference paper, attended and delivered an oral presentation in 1 international forum. Besides, two journal papers are under review.

Journals

1. **Yujiao Wu**, Liu T, Ling S H, SW Su, et al. Air Quality Monitoring for Vulnerable Groups in Residential Environments Using a Multiple Hazard Gas Detector[J]. Sensors, 2019, 19(2): 362.
2. **Yujiao Wu**, SW Su, et al. Estimation of Parameters in Electronic nose. (Under review)
3. **Yujiao Wu**, Ling S H, SW Su, et al. 3DResSA: A Novel Multimodal Deep Learning Framework for Lung Cancer Survival Analysis. (Under review)

Conference Proceedings

1. **Yujiao Wu**, Liu T, Ling S H, SW Su, et al. A Smart "E-Nose" System for Indoor Hazardous Air Monitoring. Australian Biomedical Engineering Conference(ABEC) 2018.
2. **Yujiao Wu**, SW Su, et al. A rapid disease diagnosis system based on electronic nose. China-New Zealand Investment Forum on Innovative Technologies (IFIT) 2018.

Contents

List of Figures

List of Tables

Abbreviation

2D	Two dimensions or two-dimensional
3D	Three dimensions or three-dimensional
4D	Four dimensions or four-dimensional
AI	Artificial Intelligence
ML	Machine Learning
DL	Deep Learning
CNN	Convolutional Neural Network
NN	Neural network
C-index	Concordance
MAE	Mean absolute error
MSE	Mean square error
BS	Brier score
IBS	Integrated Brier score
LC	Lung cancer
NSCLC	Non-small cell lung cancer
SCLC	Small cell lung cancer
ROC	The receiver operating characteristic curve
AUC	The area under ROC curve

kNN	K-nearest neighbors
SVM	Support vector machines
CT	Computed tomography
CAT	Computed axial tomography
LDCT	Low-dose CAT scan or CT scan
ECG	Electrocardiogram
EEG	Electroencephalogram
MRI	Magnetic resonance imaging
PET	Positron emission tomography
WHO	World Health Organization
CA	Classical adenocarcinoma
SRCC	Signet-ring cell carcinoma
MAC	Mucinous adenocarcinoma
SCC	Squamous cell carcinoma, or known as epidermoid carcinoma
QOL	Quality of life
MHGD	Multiple hazard gas detector
RDDS	Rapid disease detection system
DeepMMSA	Multimodal deep learning framework for survival analysis
TCIA	The Cancer Imaging Archive
DC	Direct current
BRI	Building-related illnesses
SBS	Sick building syndromes
IAQ	Indoor air quality

SIDS	Sudden infant death syndrome
PGA	Programmable gain amplifier
S/N	Signal-to-noise ratio
PUFA	Polyunsaturated fatty acids
ROS	Reactive oxygen species
VOC	Volatile organic compound
MOX	Metal oxide
MOS	Metal-oxide-semiconductor
QCM	Quartz crystal microbalance
BAW	Bulk acoustic wave
SAW	Surface acoustic wave
CB	Catalytic bead
TBC	Total bacterial count
MCU	Micro control unit
USB	Universal serial bus
UART	Universal asynchronous receiver transmitter
TBARS	Thiobarbituric acid reactive substances assay
TVBN	Total volatile basic nitrogen
MM	Malignant mesothelioma
ARD	Asbestos-related disease
COPD	Chronic obstructive pulmonary disease
ECOPD	An exacerbation of Chronic obstructive pulmonary disease
BN	Batch normalization

COVID-19	Coronavirus disease 2019

Notations

r_i	i-th 3D radiology image
c_i	i-th clinical information
X_i	i-th data sample, usually a vector
T_i	Time i-th event of interest or censoring occurred
$\hat{T}_i$	predicted or estimated T_i
e_i	i-th event, 1 for uncensored, 0 for censored
$I(x)$	1 if $x = True$ else 0
$S(t)$	Survival function
$\hat{S}(t)$	Estimated survival function
$h(t)$	Hazard function
$H(t)$	Cumulative hazard function
$\hat{G}(t)$	Kaplan-Meier estimator of the censoring distribution
t_{max}	Maximal time for the estimated survival time
$\sigma(x)$	Sigmoid funciton, $\frac{1}{1+exp(-x)}$
R_L	Load resistance
R_s	Sensor resistance
V_c	Circuit voltage
V_{out}	Output voltage

τ	Time constant
k	DC gain
θ	Time delay
$u(t)$	Input in the time domain
$y(t)$	Input in the time domain
$\dot{y}(t)$	The first-order form of output in the time domain
$G(s)$	Transfer function, ratio of output and input in the Laplace domain

Abstract

Lung Cancer (LC) or lung carcinoma is the uncontrolled growth of epithelial cells that line up in the respiratory tract. LC is a leading cause of cancer death in both males and females and it has contributed the deaths of millions of people around the world. Smoking tobacco products is the major cause of LC.

Tremendous progress has been made in terms of better diagnosis and treatment of LC. However, the majority of LC deaths are caused by the slow spread and development of the disease. Indeed, most LC patients are diagnosed at an advanced stage only after they have presented with obvious symptoms. Owing to this, curative treatment is no longer an option. Thus, developing effective screening methods and accurate prognosis of LC is of paramount importance, not simply for early detection but also to improve patients' quality of life (QOL) and reduce the mortality of LC. Regular chest x-rays have been studied for LC screening, but they have been of limited assistance in prolonging the lives of most patients. In recent years, low-dose CAT scan (formerly known as computed axial tomography or CAT scan) or CT scan (LDCT) has been applied to patients with a higher risk of getting LC. Nevertheless, this kind of screening needs to be conducted by appropriate CT scanners. Besides, these facilities also need to work together with the staff with rich experience in CT scans for LC screening. What's more important, to guarantee a timely treatment, a

team of specialists need to cooperate with the facilities to give patients proper health care and follow-up if there are abnormalities found. To sum up, the existing system have different kinds of deficiencies and hinders the patients of LC from getting timely treatment. Moreover, research shows that 2/3 of the world population, which equals 4.7 billion people, lacks adequate radiology specialists and the right kind of medical facilities.

To alleviate the above-mentioned issues, we propose a cheaper, easy-to-use, portable electronic nose (e-nose) system to analyze the biomarkers in human breath to rapidly and non-invasively discriminate LC patients from healthy individuals. The e-nose for volatile organic compounds (VOCs) pattering is cheaper and portable. Using cross-reactive, it can detect and discriminate between complex mixtures. To explore the working principle and discrimination ability of the e-nose system and overcome the limitations of using existing non-intelligent, slow-responding, deficient gas sensors, we proposed a novel artificial-intelligent-based multiple hazard gas detector (MHGD) system that is mounted on a motor vehicle-based robot that can be remotely controlled. First, we optimized the sensor array for the classification of three hazardous gases. After that, the optimal sensor array was mounted on the MHGD to detect and classify the target gases. Finally, MHGD is tested through experiments and the results shows that the designed MHGD system could achieve an acceptable accuracy (70.00%).

Even though the previously mentioned prototype achieved an acceptable performance for hazard mixtures classification, but we needed for medical applications a scalable, stable, and robust device with a sealable gas path and automatic control system. Thus, to enhance the system's robustness and overcome the many deficiencies in MHGD, a prototype, namely 'Rapid Disease Detection System (RDDS),' was developed. This device is designed for breath analysis in the medical field. In real-world clinical practice, multiple devices based on identical designs will be used in different clinics. Thus, it is essential to perform instrument calibration before the sampling procedure to ensure the data is reproducible and reliable for analysis. In the RDDS system, with time delays, three parameters need to be determined in the calibration process: the time delay, the gain, and the time constant. Based on this, a parameter estimation method for the RDDS system is proposed. We analyzed four different standard gas mixtures(CSGMs) to calibrate the RDDS system. Finally, we obtained the three parameters of the system with the average value for the fit to the estimation data of 92.8%.

Moreover, for a better-individualised prognosis for LC and improved survival predictability, we worked at a deeper level towards survival analysis. To reveal the underlying relation of prognostic information of radiomic images, fully utilise the potential of the prognostic power existing in the radiomic data, and exploit the correlations between radiomic images and survival information, we made the first attempt to develop a deep 3D multimodal deep learning framework for survival analysis (namely DeepMMSA) using the medical image in radiology. Quantitative results on the Non-Small Cell Lung Cancer Radiomics (NSCLC-Radiomics) data show that the proposed method could surpass the traditional methods by 4% on concordance, revealing that our method could provide a more accurate diagnosis method and prognostic decision-making solution in future clinical practice.