

國立臺灣大學醫學院暨工學院醫學工程學研究所

碩士論文

Graduate Institute of Biomedical Engineering

College of Medicine and College of Engineering

National Taiwan University

Master Thesis



肺部電腦斷層掃描之毛玻璃狀肺結節侵犯性區別診斷
:低實質比、深度學習與放射體學方法

Classification of the Invasiveness of Ground Glass Nodule
on CT images: low solid ratio, deep learning and
radiomics method

莊靜佳

Ching-Chia Chuang

指導教授：陳中明 博士

Advisor: Chung-Ming Chen, Ph.D

中華民國 111 年 6 月

June, 2022



摘要

根據衛生福利部於 109 年國人死因統計結果[1]，癌症的死亡率在十大死因中排序第一，而又以氣管、支氣管、肺癌位於癌症中的第一且已蟬聯多年，是一項不可輕忽的疾病，若能及早給予患者相應的治療的話，可以相對提高存活率。較早期的肺部腫瘤在電腦斷層掃描上通常會呈現毛玻璃狀，毛玻璃狀定義即為腫瘤呈現霧狀並且不會遮擋住穿越之血管或是蓋住氣管等組織。另外，於 2011 年有三個協會：國際肺癌協會（IASLC）、美國胸科協會（ATS）及歐洲呼吸協會（ERS），共同從病理角度針對毛玻璃狀肺腺癌做分類，以他們對人體所造成的威脅程度可以分成三大類，分別為較不具威脅性的 pre-invasive lesions，包含非典型腺瘤樣增生（AAH）與原位肺腺癌（AIS），以及另外兩類：微浸潤腺癌（MIA）和浸潤腺癌（IA）[2]，而不同分類的肺腺癌所需的開刀方式與存活率也不盡相同，其中 AIS/MIA 有些醫生的準則會選擇先以追蹤為主，或使用次肺葉切除術（楔狀切除術或是肺節切除術），在術後五年的存活率是接近 100%，反觀浸潤腺癌（IA）的存活率則有所降低（依照不同的亞型而有不同的存活率）[3]。所以若能從電腦斷層掃描上將早期毛玻璃狀腫瘤中的原位肺腺癌與微浸潤線癌從浸潤腺癌中分出來可以給予醫生應進行手術還是應先觀察的一個參考，是相當重要的事。

於本論文，將採用實質比小於 0.25 且小於三公分的毛玻璃狀原位肺腺癌（AIS）、微浸潤線癌（MIA）與浸潤腺癌（IA）當作研究材料，經醫師判斷後其須符合腫瘤實質區域之最大直徑與腫瘤本體之最大徑比值小於 0.25，並將 AIS 與 MIA 歸類為不具侵犯性的肺腺癌，而 IA 屬於具侵犯性的肺腺癌。研究目的為利用放射體學與深度學習的方法將具侵犯性的肺腺癌與不具侵犯性的肺腺癌給區分出來。

關鍵詞：肺部電腦斷層掃描影像、毛玻璃樣腫瘤、低實質比、毛玻璃樣腫瘤是否具侵犯性的分類、放射體學與深度學習



Abstract

According to the report from Ministry of the health and welfare in 2020[1], cancer ranked first in the cause of death statistics. Among all cancer cause of death, cancers of trachea, bronchus and lung were placed first for a long time. Apparently, it's important to have an early detection in order to lead to cure and enhance the survival rate since lung cancer is a force to be reckoned with. Early-stage lung adenocarcinoma nodules often manifest as ground-glass opacity (GGO) which is defined as lesions showing hazy, increased attenuation that does not obscure underlying bronchial structures or pulmonary vessels. In 2011, the International Association for the Study of Lung Cancer (IASLC), the American Thoracic Society(ATS), and the European Respiratory Society (ERS) classified lung adenocarcinomas manifest as GGO into three groups of type: (1) pre-invasive lesions, including atypical adenomatous hyperplasias (AAH) and adenocarcinoma on situ (AIS), (2) minimally invasive adenocarcinoma (MIA), and (3) invasive adenocarcinoma (IA)[2]. The lung adenocarcinomas from each groups are suggested for different therapeutic strategy. AIS and MIA can be tracked at first or treated with sublobar resection (wedge or segmental resection) with a 100% or nearly 100% of 5-year survival-rate[3]. On the other hand, the invasive adenocarcinoma causes a reduction in survival-rate (Which depends on the subtype of the adenocarcinoma). As



stated above, classifying AIS and MIA from IA which manifest as GGO on computed tomography is crucial that either gives the doctors an option to track first or to perform the operation.

In this study, the inclusion criteria are the maximum diameter is less than 3 cm and the solid ratio of the ground-glass nodules must be less than 0.25 which judged by the doctor, indicates the ratio of max diameter of the solid lesion to max diameter of the whole lesion need to be less than 0.25. Furthermore, AIS and MIA are classified as non-invasive adenocarcinoma while IAs are invasive adenocarcinoma. The purpose of the study is to use radiomics and deep learning to build up classification model so as to make a precise precision to the classification problem.

Keywords: Lung computed tomography scan, ground glass nodule ,low solid ratio, classification of the invasiveness of the ground glass nodules, radiomics and deep learning



目錄

摘要	II
Abstract	III
圖目錄	VI
表目錄	VIII
第一章 緒論	1
1.1 研究背景	1
1.2 肺癌分類與分期介紹	3
1.3 AIS, MIA, IA 定義	5
1.4 研究動機與目的	8
第二章 文獻探討	13
2.1 毛玻璃狀腫瘤之侵犯性分類—放射體學模型	13
2.2 毛玻璃狀腫瘤之侵犯性分類—深度學習	17
2.3 毛玻璃狀腫瘤之侵犯性分類—放射體學結合深度學習	18
第三章 研究方法與材料	21
3.1 研究材料	21
3.2 研究方法	22
3.2.1. Adaboost 演算法	23
3.2.2. 影像處理	24
3.2.3. 模型架構	28
3.2.4. 深度模型更新資料與權重	37
3.3 效能評估	41
3.3.1. 效能指標	41
3.3.2. 比較文獻介紹之方法	44
3.3.3. 本研究方法之比較	65
第四章 研究結果與討論	77
4.1 針對分不好的資料進行著重學習的三種方式	77
4.2 kernel number 數量的影響	78
4.3 Attention layer 的重要性	79
4.4 影像輸入	80
4.5 fusion	81
4.6 Solid ratio 的影響	82
第五章 結論與未來展望	85
5.1 結論與未來展望	85
5.2 研究限制	86
參考資料	87



圖 目 錄

圖 1.1 109 年十大死因死亡人數[1].....	1
圖 1.2 109 年男女前十大死因[1].....	2
圖 1.3 109 年前十大癌症死因[1].....	2
圖 1.4 AIS 病理圖[7].....	6
圖 1.5 MIA 病理圖[7]	7
圖 1.6 IA 病理圖[7].....	8
圖 1.7 毛玻璃狀腫瘤 CT 圖	9
圖 1.8 實質化腫瘤 CT 圖	9
圖 2.1 Residual learning-based CNN network [12].....	18
圖 2.2 Flowchart[13].....	19
圖 3.1 最大直徑與 mean CT value 分布圖	22
圖 3.2 Adaboost 演算法流程圖	24
圖 3.3 原始影像圖	26
圖 3.4 圖 3.3 之分割圖	27
圖 3.5 Convolution 卷積運算 (a)是原始影像 4*4 , (b)為 kernel , (c)為原始影像做完卷積之結果	30
圖 3.6 Depth-wise Convolution[18]	31
圖 3.7 Maxpooling 示意圖	31
圖 3.8 Global Average Pooling	32
圖 3.9 Fully connected layer	33
圖 3.10 relu	34
圖 3.11 Batch Normalize 與 Activation layer	34
圖 3.12 SE net[22]	35
圖 3.13 毛玻璃狀腫瘤 CT 影像	36
圖 3.14 AdaBoost model.....	36
圖 3.15 Adaboost Model	37
圖 3.16 AdaBoost CNN Flowchart[23]	39
圖 3.17 AdaBoost Model 抽取樣本流程圖	41
圖 3.18 混淆矩陣	42
圖 3.19 5-fold cross validation	44
圖 3.20 本研究資料重複 Fan Li [8]等人提倡方法之 ROC curve	45
圖 3.21 Fan Li [8]等人提倡方法之 ROC curve	45



圖 3.22 model 1 產生之結果	48
圖 3.23 model 2 產生之結果	48
圖 3.24 文獻 gross radimoics score 於訓練集與測試集之箱型圖[10]	50
圖 3.25 (a)~(e)為利用腫瘤內部之特徵重複文獻做法其 Fold 1 至 Fold 5 分別在做完 SMOTE 之 training data 兩類分布箱型圖、原始 training data 兩類分布箱型 圖、testing data 兩類分布箱型圖(f)為 5-fold 之 ROC.....	52
圖 3.26 (a)~(e)為利用腫瘤外部之特徵重複文獻做法其 Fold 1 至 Fold 5 分別在做完 SMOTE 之 training data 兩類分布箱型圖、原始 training data 兩類分布箱型 圖、testing data 兩類分布箱型圖(f)為 5-fold 之 ROC.....	54
圖 3.27 (a)~(e)為利用腫瘤外部與內部之特徵重複文獻做法其 Fold 1 至 Fold 5 分別 在做完 SMOTE 之 training data 兩類分布箱型圖、原始 training data 兩類分布 箱型圖、testing data 兩類分布箱型圖(f)為 5-fold 之 ROC	56
圖 3.28 模型與醫師判斷之 ROC 曲線[11]	60
圖 3.29 本研究測試資料之 ROC 曲線	60
圖 3.30 Radiomics 模型與深度學習模型以及 fusion model 之 ROC 曲線[13]	61
圖 3.31 本研究資料 Radiomics 模型與深度學習模型以及 fusion model 之 ROC 曲線	62
圖 3.32 (a) testing 以及 validation mean CT value (b) testing data mean CT value 箱型 圖[13]	63
圖 3.33 本研究資料 mean CT value 箱型圖	63
圖 3.34 調整過後的三個模型 ROC 曲線.....	64
圖 3.35 Combine 的方式，本圖為示意圖，僅呈現其中一個 convolution kernel 之 combine 方式，在其他三個 convolution kernel 會用一樣的方式進行結合	66
圖 3.36 Average 的方式，本圖為示意圖，僅呈現其中一個 convolution kernel 之 average 方式，在其他三個 convolution kernel 會用一樣的方式進行權重平均	67
圖 3.37 geneticCNN	67
圖 3.38 結合模型 dense 層特徵做特徵挑選與分類	70
圖 3.39 pretrain 及 finetune 簡易流程圖	71
圖 3.40 Ensemble 模型架構(a) model 1 (b) model 2.....	72
圖 3.41 (a) Atrous convolution (b) Atrous spatial pyramid pooling.....	73
圖 3.42 Residual block.....	73
圖 3.43 fireblock.....	74
圖 3.44 每回合進行之模型訓練圖	75
圖 4.1 Attention layer 的位置	80
圖 4.2 原始 VOI (上排) 與只有腫瘤部分之 VOI (下排)	81
圖 4.3 輸入影像為原始影像之深度、radiomics、fusion 模型 ROC 曲線	82
圖 4.4 mean CT value histogram.....	83



表目錄

表 1.1 TNM 描述子定義[6].....	3
表 1.2 肺癌分期(第八版) [6].....	5
表 1.3 AIS 定義[7].....	6
表 1.4 MIA 定義[7]	7
表 1.5 小於三公分之毛玻璃狀腫瘤在影像與病理角度之分類[7].....	8
 表 3. 1 兩類資料之 CT 參數與最大直徑、mean CT value 比較表	22
表 3.2 每個 fold 各自的訓練、驗證、測試及資料數量	44
表 3.3 Fan Li [8] 等人研究資料 mean CT value 比較	46
表 3.4 本研究資料之 mean CT value	46
表 3.5 文獻結果統計表[9]	47
表 3.6 文獻對於訓練及與驗證集資料比較表[9].....	49
表 3.7 本研究資料於兩類別之最大直徑分布	49
表 3.8 文獻不同模型結果比較[11]	57
表 3.9 本研究資料重現文獻之方法結果.....	58
表 3.10 本研究資料與文獻資料於 solid portion 及 attenuation 統整表	58
表 3.11 放射科醫師對 GGN 分類標準[12].....	60
表 3.12 文獻結果與本研究重現文獻做法之結果統計表.....	64
表 3.13 平均權重之效能統計表	68
表 3.14 結合權重之效能統計表	68
表 3.15 模型輸入之 feature map	74
表 3.16 原始模型與多 feature map 輸入之結果比較.....	76
 表 4.1 分類較不好的樣本三種著重學習的方式之 AUC 結果.....	78
表 4.2 方法三最佳方法之正確率、靈敏度、特異度.....	78
表 4.3 不同 kernel number 之 AUC 結果	79
表 4.4 原始 VOI 與僅有腫瘤部分之 VOI 效果比較圖	81
表 4.5 三種資料之 mean CT value 統整表.....	83
表 4.6 三種資料之 IA、NonIA 個數比較.....	83
表 4.7 三種資料之平均 AUC.....	84



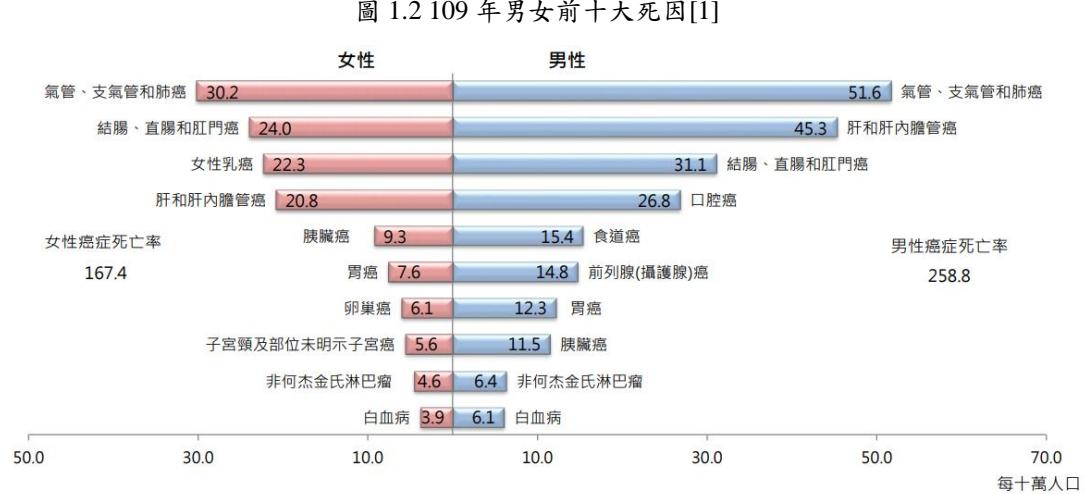
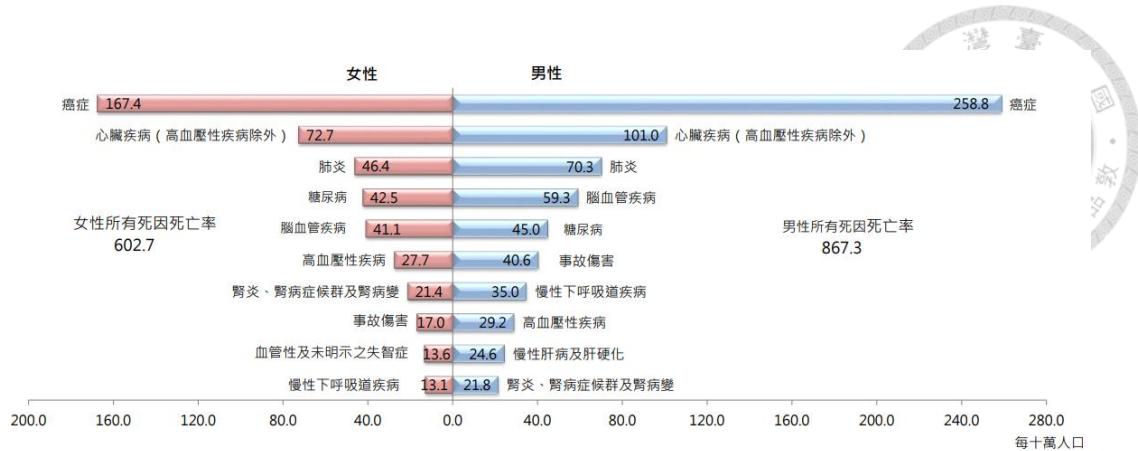
第一章緒論

1.1 研究背景

根據衛生福利部於 109 年國人死因統計結果[1]，癌症的死亡率在十大死因同 108 年之排名為排序第一，其數量遠高於剩餘的死亡原因（與第二名心臟疾病相差近三萬人，如圖 1.1），且不論女性或是男性皆為癌症死亡率最高，如圖 1.2 所示，其中又以氣管、支氣管、肺癌位於癌症中的第一並已蟬聯 41 年之多（見圖 1.3），是一項長期為國人所惱的疾病，若能及早得到相應治療的話，便可以相對提高患者的存活率[4]。

順位	所有死亡原因	死亡人數(人)		死亡率 (每十萬人口)			標準化死亡率 (每十萬人口)		
		109年	較上年 增減%	108年 順位	109年	較上年 增減%	順位	109年	較上年 增減%
1	癌症	50,161	-0.1	1	212.7	-0.1	1	117.3	-3.3
2	心臟疾病 (高血壓性疾病除外)	20,457	3.0	2	86.7	3.1	2	43.8	0.4
3	肺炎	13,736	-9.5	3	58.2	-9.5	3	26.4	-12.1
4	腦血管疾病	11,821	-2.9	4	50.1	-2.9	4	25.2	-5.4
5	糖尿病	10,311	3.2	5	43.7	3.2	5	22.0	-1.1
6	事故傷害	6,767	1.9	6	28.7	2.0	6	20.3	1.1
7	高血壓性疾病	6,706	7.2	8	28.4	7.3	8	13.4	3.9
8	慢性下呼吸道疾病	5,657	-10.2	7	24.0	-10.2	7	11.0	-12.6
9	腎炎、腎病症候群及腎病變	5,096	0.9	9	21.6	1.0	9	10.5	-2.3
10	慢性肝病及肝硬化	3,964	-6.5	10	16.8	-6.5	10	10.3	-7.8

圖 1.1 109 年十大死因死亡人數[1]



肺癌死亡率之所以這麼高有很大的原因是早期肺腺癌腫瘤分佈於肺部周邊，不會造成明顯的症狀，以致於通常發現的時候都已較晚期，癌細胞增生的很快，導致腫瘤早已開始擴散，即使發現了再進行手術切除，對於這種較晚期的肺腫瘤復發率仍舊很高[5]。由於手術治療方式對於肺腺癌的預後非常重要，而肺腺癌的分期和是否能接受手術與手術方式也有密切的關係，所以如果可以提早透過電腦斷層掃瞄影像精確判斷早期的毛玻璃狀腫瘤是否屬於較侵犯性的腫瘤，就可以提供醫生一個判斷的參考依據。近幾年也開始有相關文獻對這方面的研究做探討，不論是根據臨床方面找出的特徵對其做區別（利用腫瘤的型態、病人的生理特徵、習慣等等...）或是利用工程的方式去對腫瘤本身找特徵做分類（包含放射體學亦或是深度學習等等...）皆可找到相關文獻，尤以近幾年所發表出的相關文獻居多，可見其重要性不容小覷且越來越受到重視。



1.2 肺癌分類與分期介紹

肺癌可以分為兩大類：非小細胞癌（佔約 85%）與小細胞癌（佔約 15%），其中非小細胞癌是這次研究針對的目標群。International Association for the Study of Lung Cancer (IASLC) 公布新修訂的肺癌分期是根據最新 TNM 分類版本—2017 年一月由 American Joint Commission on Cancer (AJCC) 及 Union for International Cancer Control (UICC) 發表出來的第八版[6]所制定，見表 1.1，第八版依據 2010 年修訂的肺癌國際分期法（第七版），再度對非小細胞癌的分期做更進一步的調整，並於 2018 年一月開始實施。TNM 分期中的 primary tumor (T)：是根據原始腫瘤特性去做分類（從輕微到最嚴重分別為：T0, Tis, T1, T2, T3, T4）、regional lymph nodes (N)：則是判斷是否有區域性的淋巴結侵犯（侵犯區域的不同給予不一樣的分級：N0, N1, N2, N3）、最後 distant metastasis (M)：觀察是否有遠端轉移（不同程度一樣給予不同分級：M0, M1a, M1b, M1c），最後再統整三者 primary tumor(T)、regional lymph nodes(N)、distant metastasis(M)，依照不同的嚴重性判定肺癌的分期：(由輕微到嚴重) IA1, IA2, IA3, IIA, IIIA, IB, IIB, IIIB, IIIC, IVA, IVB，如表 1.2 所示。

表 1.1 TNM 描述子定義[6]

TNM 描述子定義 -Definition for TNM descriptors

T (primary tumor)	
T0	No primary tumor
Tis	Carcinoma in situ (squamous or adenocarcinoma)
T1	Tumor ≤ 3 cm
T1mi	Minimally invasive adenocarcinoma
T1a	Superficial spreading tumor in central airways*
T1a	Tumor ≤ 1 cm



T1b	Tumor > 1 but ≤ 2 cm
T1c	Tumor > 2 but ≤ 3 cm
T2	Tumor > 3 but ≤ 5 cm or tumor involving: visceral pleura, [†] main bronchus (not carina), atelectasis to hilum [†]
T2a	Tumor > 3 but ≤ 4 cm
T2b	Tumor > 4 but ≤ 5 cm
T3	Tumor > 5 but ≤ 7 cm or invading chest wall, pericardium, phrenic nerve; or separate tumor nodule(s) in the same lobe
T4	Tumor > 7 cm or tumor invading: mediastinum, diaphragm, heart, great vessels, recurrent laryngeal nerve, carina, trachea, esophagus, spine; or tumor nodule(s) in a different ipsilateral lobe

N(regional lymph nodes)

N0	No regional node metastasis
N1	Metastasis in ipsilateral pulmonary or hilar nodes
N2	Metastasis in ipsilateral mediastinal or subcarinal nodes
N3	Metastasis in contralateral mediastinal, hilar, or supraclavicular nodes

M (distant metastasis)

M0	No distant metastasis
M1a	Malignant pleural or pericardial effusion [‡] or pleural or pericardial nodules or separate tumor nodule(s) in a contralateral lobe
M1b	Single extrathoracic metastasis
M1c	Multiple extrathoracic metastases (1 or>1 organ)

*Superficial spreading tumor of any size but confined to the tracheal or bronchial wall. †Atelectasis or obstructive pneumonitis extending to hilum; such tumors are classified as T2a if >3 and ≤ 4 cm, T2b if > 4 and ≤ 5 cm. ‡Pleural effusions are excluded that are cytologically negative, nonbloody, transudative, and clinically judged not to be due to cancer.



表 1.2 肺癌分期(第八版) [6]

肺癌分期(Lung Cancer Stage Grouping)

T / M	Subcategory	N0	N1	N2	N3
T1	T1a	IA1	IIB	IIIA	IIIB
	T1b	IA2	IIB	IIIA	IIIB
	T1c	IA3	IIB	IIIA	IIIB
T2	T2a	IB	IIB	IIIA	IIIB
	T2b	IIA	IIB	IIIA	IIIB
T3	T3	IIB	IIIA	IIIB	IIIC
T4	T4	IIIA	IIIA	IIIB	IIIC
M1	M1a	IVA	IVA	IVA	IVA
	M1b	IVA	IVA	IVA	IVA
	M1c	IVB	IVB	IVB	IVB

1.3 AIS, MIA, IA 定義

本研究所針對之肺腫瘤類型為非小細胞癌中的肺腺癌，並且其於 primary tumor (T) 的等級為介於 Tis 到 T1 之間的原位肺腺癌 (adenocarcinoma on situ , AIS)、微浸潤腺癌 (minimally invasive adenocarcinoma , MIA) 及浸潤腺癌 (invasive adenocarcinoma , IA)，這個階段的肺腺癌通常都是較為早期、呈現毛玻璃狀的腫瘤。在 IASLC/ATS/ERS 中分別對 AIS, MIA, IA 有各自的定義[7]，且自 2015 年 WHO 即開始使用此定義，其中 AIS (表 1.3、圖 1.4) 指的是最長徑小於 3 公分的肺腺癌，並只由 lepidic 組成，裡面沒有任何 invasive component pattern (像是 acinar, papillary...)、沒有任何侵犯性特徵像是基質、血管、肋膜侵犯，也不會有腫瘤經肺泡間隙擴散 (spread through air spaces, STAS)，另外其細胞型態通常是非黏性的 (nonmuconous)，但有時還是會有例外。至於 MIA (表



1.4、圖 1.5)，大部分的定義都與 AIS 相同，他們之間不同的點在於構成 MIA 之「主要」組成成分是 lepidic，所以不一定會是整顆都由 lepidic 組成，可以接受小於等於 0.5 公分的 invasive component（像是其他的亞型 acinar, papillary, micropapillary, solid, colloid, ...），除了以上兩者的定義，其餘就屬於 IA（圖 1.6）。

表 1.3 AIS 定義[7]

原位癌 Adenocarcinoma in situ (AIS)

Pathologic criteria

- A small tumor ≤ 3 cm
 - A solitary adenocarcinoma
 - Pure lepidic growth
 - No stromal, vascular, or pleural invasion
 - No pattern of invasive adenocarcinoma (such as acinar, papillary, micropapillary, solid, colloid, enteric, fetal, or invasive mucinous adenocarcinoma)
 - No spread through air spaces
 - Cell type mostly nonmucinous (type II pneumocytes or Clara cells), rarely may be mucinous (tall columnar cells with basal nuclei and abundant cytoplasmic mucin, sometimes resembling goblet cells)
 - Nuclear atypia is absent or inconspicuous
 - Septal widening with sclerosis/elastosis is common, particularly in nonmucinous adenocarcinoma in situ
-

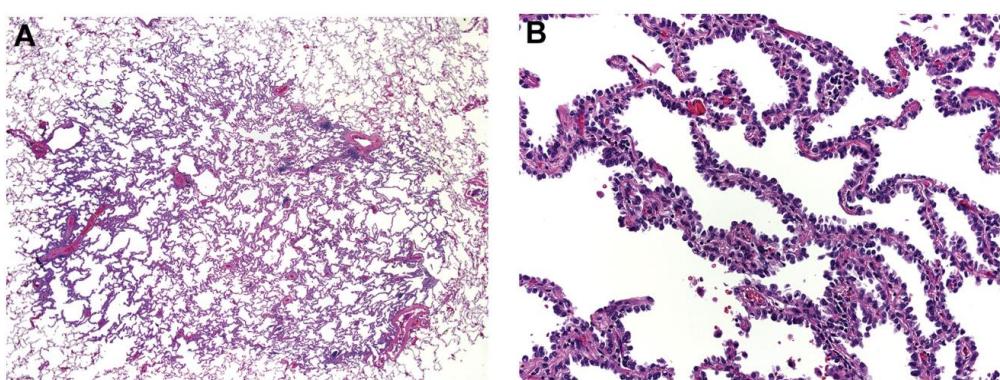


圖 1.4 AIS 病理圖[7]

表 1.4 MIA 定義[7]



Minimally Invasive Adenocarcinoma (MIA)

Pathologic criteria

- A small tumor 3 cm
- A solitary adenocarcinoma
- Predominantly lepidic growth
- Invasive component 0.5 cm in greatest dimension in any one focus
- Invasive component to be measured includes
 1. Any histologic subtype other than a lepidic pattern (such as acinar, papillary, micropapillary, solid, colloid, fetal, or invasive mucinous adenocarcinoma)
 2. Tumor cells infiltrating myofibroblastic stroma
- The diagnosis of minimally invasive adenocarcinoma is excluded if the tumor
 1. Invades lymphatics, blood vessels, air spaces, or pleura
 2. Contains tumor necrosis,
 3. Spread through air spaces
- The cell type in most cases consists of nonmucinous (type II pneumocytes or Clara cells), but rarely may be mucinous (tall columnar cells with basal nuclei and abundant cytoplasmic mucin, sometimes resembling goblet cells)

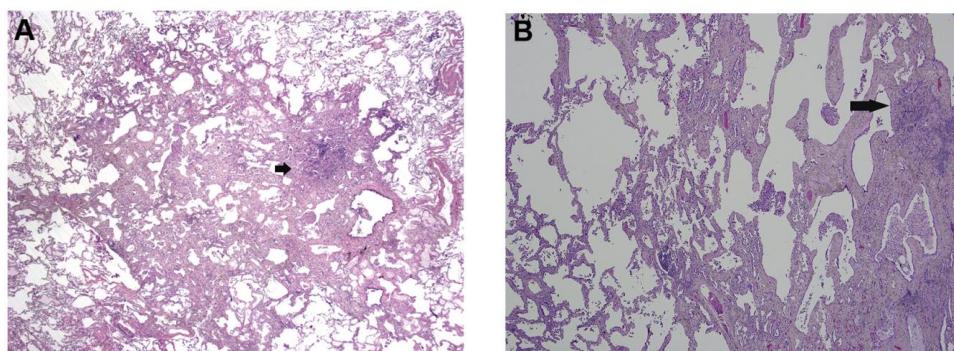


圖 1.5 MIA 病理圖[7]

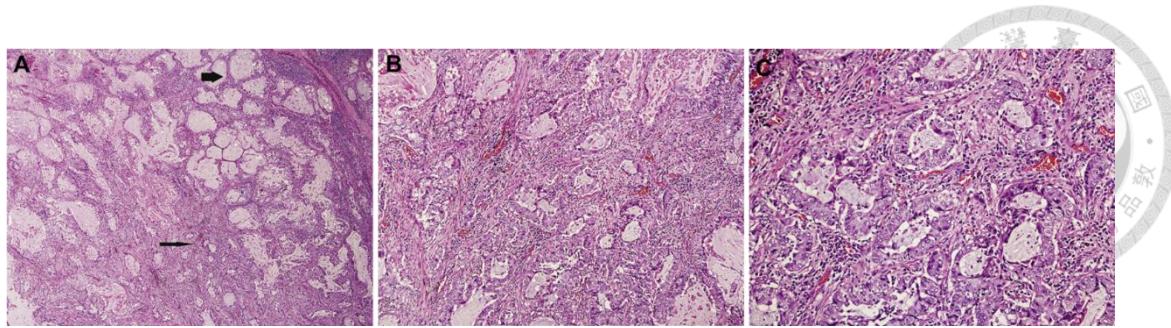


圖 1.6 IA 病理圖[7]

表 1.5 為這次第八版中對小於三公分且含有毛玻璃狀及 lepidic component 的肺腺癌從臨床（clinical, cT）及病理角度（pathologic, pT）所做的統整表，裡面有 AIS、MIA、LPA 各別在兩種角度對於整體腫瘤與 solid part 之間的比例以及他在臨床分期各屬於甚麼（兩者分期皆相同）。

表 1.5 小於三公分之毛玻璃狀腫瘤在影像與病理角度之分類[7]

	CT image on HRCT						
cT*	Solid part	0 cm	0 cm	≤ 0.5 cm †	0.6-1.0 cm †	1.1-2.0 cm †	2.1-3.0 cm †
	Total tumor size including GG	≤ 0.5 cm	0.6-3.0 cm ‡‡	≤ 3 cm ‡‡	0.6-3.0 cm ††	1.1-3.0 cm ††	2.1-3.0 cm ††
	Pathologic Differential Diagnosis	AAH‡, AIS, MIA	AIS, MIA, LPA	MIA, LPA, AIS	LPA, Invasive AD, MIA	LPA, Invasive AD	Invasive AD
	Clinical Stage*		cTis ‡‡	cT1mi‡‡	cT1a	cT1b	cT1c
pT	Invasive part	0 cm	0 cm	≤ 0.5 cm‡‡	0.6-1.0 cm†	1.1-2.0 cm†	2.1-3.0 cm†
	Total tumor size including lepidic growth part	Usually ≤ 0.5 cm‡	≤ 3.0 cm‡‡	≤ 3.0 cm‡‡	0.6-3.0 cm††	1.1-3.0 cm††	2.1-3.0 cm††
	Pathology	AAH	AIS	MIA	Lepidic predominant AD or Invasive AD with lepidic component	Invasive AD with a lepidic component or lepidic predominant AD	Invasive AD with lepidic component
	Pathologic Stage		pTis‡‡	pT1mi‡‡	pT1a	pT1b	pT1c

1.4 研究動機與目的

由於肺癌對人體造成的威脅不容小覷，所以盡早在腫瘤還是早期的時候做適



當的處理為佳。較早期的肺部腫瘤在電腦斷層掃描上通常會呈現毛玻璃狀，毛玻璃狀定義即為腫瘤呈現霧狀並且不會遮擋住穿越的血管或是蓋住氣管等組織，這種毛玻璃狀的肺腫瘤有可能屬於比較不具有威脅性的非典型腺瘤樣增生（AAH）、原位肺腺癌（AIS）、微浸潤腺癌（MIA）或者是浸潤腺癌（IA）（毛玻璃狀腫瘤 CT 圖見圖 1.7，實質化腫瘤 CT 圖見圖 1.8），這四者於 2011 年，由三個協會：國際肺癌協會（IASLC）、美國胸科協會（ATS）及歐洲呼吸協會（ERS），共同從病理角度針對肺腺癌分成三大類，分別為 pre-invasive lesions，包含非典型腺瘤樣增生（AAH）與原位肺腺癌（AIS）、微浸潤腺癌（MIA）、與較具侵犯性的浸潤腺癌（IA）[2]，而不同分類的肺腺癌所需的開刀方式與存活率也不盡相同，其中 AIS/MIA 通常會先進行追蹤觀察或是使用次肺葉切除術（楔狀切除術或是肺節切除術），在術後五年的存活率是接近 100%，反觀浸潤腺癌（IA）的存活率則降低不少（依照不同的亞型而有不同的存活率）[3]。所以若能從電腦斷層掃描上將早期毛玻璃狀腫瘤的原位肺腺癌與微浸潤線癌從浸潤腺癌區分出來不但可以給予醫生在開刀時應採取何種切除方式當作參考，還能多增加存活率，是相當重要的事。



圖 1.7 毛玻璃狀腫瘤 CT 圖

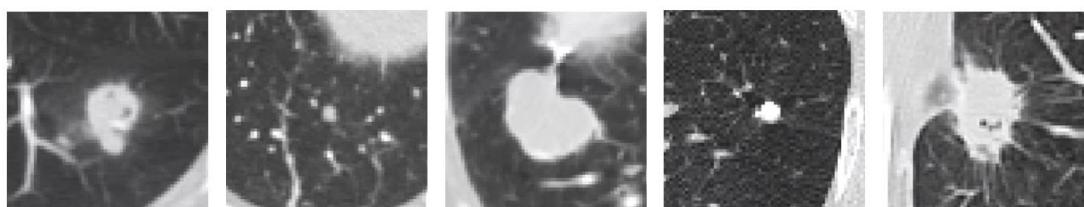


圖 1.8 實質化腫瘤 CT 圖

在目前的文獻當中，利用放射體學建立模型進行與本研究相關之分類文獻



有：2019 年，Fan Li [8] 等人利用 radiomics features 來建立分類具侵犯性與不具侵犯性腫瘤的模型，而本篇文獻與本研究略為不同的部分為其不具侵犯性的類別包含 atypical adenomatous hyperplasia (AAH) 並且並未針對 solid ratio 有特別的規範。作者透過 ICC、p-value、coefficient correlation 以及 least absolute shrinkage and selection operator (Lasso) logistic regression 的方式選出兩個代表性特徵建立模型，而這兩個特徵皆來自 GLCM，分別為 cluster tendency 與 correlation，在效能評估方面，此文獻共有三組測試集，其中一組為 intra-class validation cohort 與兩組 external cohort，三組分別的正確率為：90.8%、84.0%、88.1%，靈敏度：97.5%、82.2%、85.7%，特異度：94.4%、86.7%、89.8%，AUC：0.971、0.942、0.936。於 2020 年，Hwan-ho Cho [9] 等人，利用腫瘤最大之 2D 切面提取基本 radiomics 特徵，並另外提取出作者新發明之 marginal radiomics feature 共四個。先由基本的 radimoics features 挑選出五個特徵，分別為：Range、GLCM-entropy、ISZM-size zone variability、density、mass，再加入 marginal radiomics features 當作代表性特徵，通過 logistic regression 分類器得到的最佳正確率、靈敏度、特異度與 AUC 分別為：91.49%、89.47%、100%、0.9825，而此篇對於 part-solid GGN 的 solid 部分限制為 < 5mm。同年，Linyu Wu [10] 等人同樣針對 pure 與 part-solid GGN 做侵犯性分類，但未明確規範 part solid 腫瘤的 solid 部分占整顆腫瘤的比例。此篇文獻針對腫瘤本體與其往外膨脹 2mm 之範圍做 radiomics 特徵提取，透過 SMOTE 將兩類資料的比例調成接近 1:1，並使用 mRMR 挑選出最具相關且最小冗餘的 20 個特徵進入 lasso 得到最具代表性的特徵，將此特徵與其從 lasso 計算出來的係數做相乘後全部相加放入 logistic regression 做分類。其中以腫瘤內部與外部特徵合在一起執行上述之步驟得到的結果相較於分別執行為佳，正確率為：88.6%，靈敏度：80%，特異度：92%，及 AUC 為：0.876。同年另一篇文獻，Guangyao Wu [11] 等人也是針對 part solid 之腫瘤進行是否具侵犯性的分類，而此篇文獻有特別將 pure GGN 從資料集中挑出，以針對 part solid 之腫瘤進



行分析，另外此篇也沒有對 solid partition 進行限制，作者提出之最佳模型是透過以-50HU 當作閾值，分別產生-50HU 以上範圍（solid part）以及以下範圍（ground glass part）後，提取兩者範圍之 radiomics 特徵後進行 boruta 特徵挑選，最後透過 random forest 建構分類模型得到的結果最佳正確率、靈敏度、特異度與 AUC 分別為：93%、98%、78%、0.9800。另外，利用深度學習與同時使用深度學習和放射體學方法進行與本研究相關之分類文獻各有一篇：2020 年由 Jing Gong [12]等人提出利用深度學習的模型架構進行是否具侵犯性分類，其資料納入條件沒有針對 solid ratio 做限制，接受任何有呈現毛玻璃狀且小於三公分之腫瘤進行分類判斷。此篇文獻使用之模型是基於 residual learning-based 的 CNN network 進行修改，在 AUC 的效能上能達到 0.92 ± 0.03 的效果，並與兩位放射科醫師判斷結果作比較分別為 0.92 ± 0.03 與 0.90 ± 0.03 。於同年，Xianwu Xia[13]等人提出結合深度學習與放射體學的方法來對 non-invasive 與 invasive 腫瘤分類，文獻中有提到納入的資料為 pGGN （pure ground glass nodule）與 sGGN （part-solid ground glass nodule），其方法為先利用分割模型 U-net 將腫瘤邊界取出，接著進入下一步會有兩個部分一邊是提取邊界內之腫瘤 radiomics 特徵，並做好挑微選取後得到 20 個代表性特徵進入 SVM 進行分類，另一邊則會 tranfer 分割模型之權重至分類模型，進行分類的訓練（fine-tune），最後將兩個預測出來的機率做結合，而最佳結合方式為取其最大機率值作為最後預測結果，AUC 為 0.90 ± 0.03 ，正確率： 80.3% 。

由於 solid ratio 對於判斷是否具侵犯性有一定的判斷能力，而上述提出之文獻未特別針對 solid ratio 進行規範，與一部份文獻沒對大小有所限制以分類具侵犯性與不具侵犯性的腫瘤，所以本研究欲透過建立深度學習網路架構，來針對資料挑選上限制在 solid ratio < 0.25 及腫瘤最大直徑小於三公分之資料進行分類，希望相較於上述文章利用放射體學與深度學習的方法，能有更好的分辨效果。

於本論文，為更符合醫師們在臨床上的需求，將採用實質比小於 0.25 且小於



三公分的毛玻璃狀原位肺腺癌（AIS）、微浸潤線癌（MIA）與浸潤腺癌（IA）當作研究材料，其須符合腫瘤實質區域的最大徑與腫瘤本體之最大徑比值經醫師判斷後小於 0.25，並將 AIS 與 MIA 歸類為不具侵犯性的肺腺癌，而 IA 屬於具侵犯性的肺腺癌，實質比設為 0.25 是根據日本兩篇有名的試驗而定[14-15]。研究目的為利用放射體學與深度學習的方法將具侵犯性的肺腺癌與不具侵犯性的肺腺癌在還是呈現毛玻璃狀、較早期時就給區分出來，以利醫師們在開刀上面做選擇。



第二章文獻探討

由於近幾年來，許多研究開始對毛玻璃狀腫瘤具侵犯性與較不具侵犯性的分類議題感興趣，由此看出此類研究是有相當程度上的重要性。在現有的文獻當中，對於判斷是否具侵犯性的分類方式主要有放射體學方法、深度學習，還有融合放射體學與深度學習兩種方式的做法，或是利用臨床資訊來進行分類，此章節將針對利用放射體學與深度學習來做分類的五篇文獻進行整理與討論。

2.1 毛玻璃狀腫瘤之侵犯性分類—放射體學模型

放射體學的方法大致可以分成四個步驟，(1) 從影像中分割出想要做分析的部分，以本研究的肺部腫瘤為例，即為在電腦斷層掃描上將腫瘤部份區域分割出來，(2) 針對分割出來的區域做特徵提取，而常見的 radiomics 像是紋理特徵 grey level co-occurrence matrices (GLCM)、grey level run-length matrices (GLRLM)、grey level size-zone matrices (GLSZM) 或是一階特徵 (Histogram) …等等，都可以從圖像提取出利用特定公式量化過的腫瘤特徵，(3) 從前面所提取出的特徵挑選出真正對研究分類有鑑別力的特徵，(4) 利用前一步驟選取出來具代表性的特徵建立模型，對資料進行分類。不同的文獻除了使用之資料不相同外，提取特徵的方法、使用的特徵挑選方式 (lasso, forward selection, ...) 與最後進行決策的分類器 (SVM, Logistic regression,...) 等等也會有些許不同，這些都有可能是影響最後分類出來的結果的變因。

2019 年，Fan Li [8]等人利用 radiomics features 建立多特徵分類模型，目的為將具有侵犯性的腫瘤從不具侵犯性腫瘤中分類出來，而此文獻之資料與本研究略為不同的部分為不具有侵犯性之腫瘤包含與 AIS 同屬 pre-invasive lesion 的 atypical adenomatous hyperplasia。另外，在材料挑選方面，作者對於納入的資料沒有針對 solid ratio 的範圍做限制，僅提及樣本為呈現毛玻璃之腫瘤且沒有助



膜或胸腔黏連。文中 radiomics 的特徵都是透過 MATLAB 2014a 提取出來的，共有 355 個。挑選特徵的方式是先挑出 intraclass correlation coefficients (ICC) > 0.75 及 p-values < 0.05 的特徵，再將具有高相關性的特徵 (correlation coefficient > 0.6) 拿掉，以減少冗餘的特徵，最後利用前面挑選出的 28 個特徵透過 least absolute shrinkage and selection operator (Lasso) 挑出係數不為 0 之最具代表性的兩個特徵建立 logistic regression 模型，兩個特徵皆來自 GLCM，分別為原始影像計算出的 cluster tendency 與 x,y,z 軸皆做過 low-pass filter 之 wavelet 影像計算的 correlation，其分類迴歸式為： $2.307 + 4.995 \times \text{correlation} - 7.660 \times \text{cluster tendency}$ 。而這篇文獻僅有一組訓練集並透過三組測試集對訓練完之模型做評估，測試集分別為一組 intra-class validation cohort 及兩組 external cohort，三組得到的正確率分別為：90.8%、84.0%、88.1%，靈敏度分別為：87.5%、82.2%、85.7%，特異度則為：94.4%、86.7%、89.8%，而 AUC 分別是：0.971、0.942、0.936。

2020 年有三篇分類肺腺癌是否具侵犯性的相關文獻，首先第一篇由 Hwan-ho Cho [9]等人提出的，此篇的分類目標與本研究相同，同為想將 IA 從 MIA/AIS 區分出來，另外文獻有提及納入的資料對於 part-solid 的 solid 部分限制為 < 5mm，但對於 solid ratio 的部分並未提出其規範。作者利用一組訓練集 (189 筆) 來建構模型並用一組測試集 (47 筆) 測試三種不同分類器所建立之模型的分類結果。這篇文獻計算腫瘤特徵的範圍是針對具有最大腫瘤切面的那張電腦斷層掃描影像進行計算，為一張 2D 的影像，而提取的特徵共有 40 個，其特別的地方在於它除了使用基本的 radiomics 特徵（面積、質量、密度、灰階亮暗變化共 19 個與紋理特徵 GLCM 共 16 個、ISZM 共 2 個）外，還新增一組特徵：marginal radiomics feature，這個想法是來自於作者認為腫瘤以灰階的角度來看大致分成兩種，一種是同質性較高的腫瘤（文中稱為：well-defined tumor），當灰階分別以 2D 切面中腫瘤內部灰階值之 5/100, 35/100, 65/100, 95/100 作為計算的閾值，其產生



的這組灰階閾值從低到高攀升，並計算在每一個灰階閾值下整顆腫瘤灰階低於此閾值之 pixel 數量占整顆腫瘤 pixel 數量的比例繪製成累積分布函數 CDF，這類型腫瘤會在此 CDF 曲線下找到某兩個相鄰灰階閾值得到之比例有極大的落差，即瞬間上升，因為其大部分腫瘤內灰階都存在於兩個閾值間的範圍內；另一類異質性較高的腫瘤（文中稱為：ill-defined tumor）其產生之 CDF 則是會緩緩上升，不會有一個明顯的陡峭的落差，因為腫瘤內的灰階值分布範圍較為廣泛，所以較不會有大部分灰階值都落在某範圍內的情況。繪製完累積分布函數圖後，對這個曲線取斜率再計算出下列四個特徵：平均、標準差、峰度與偏態，即為 marginal radiomics feature。提取完特徵後，是挑選特徵的方式，這篇文章首先用 LASSO 的方式從原始的 40 個特徵裡面選出五個代表性特徵（Range, GLCM-entropy, ISZM-variability, density, mass）後加入前面提到之新特徵—marginal radiomics features，並放入三種不同的分類器對資料進行分類，看其分類效果如何。以結果最好的 logistic regression 分類器來說，正確率、靈敏度、特異度與 AUC 分別為：91.49%、89.47%、100%、0.9825，相較於尚未加入 marginal radiomics feature 的正確率、靈敏度、特異度與 AUC 分別為：89.36%、89.47%、88.89%、0.9766 有一定程度的提升。

接著第二篇，由 Linyu Wu [10] 等人撰寫出的文獻中，樣本資料納入 pure GGN 及 part-solid 的腫瘤，但沒有明確規範 part solid 中的 solid 占整顆腫瘤的比例，僅針對腫瘤大小做限制：小於三公分。此外，這篇文章同樣只分成一組訓練集及測試集資料比例為 3:7。首先，文獻著重於把腫瘤內部特徵與腫瘤外部特徵同時提取出來做特徵選取與分類，其結果對於僅提供腫瘤內特徵所分類之結果與僅用腫瘤外特徵之分類結果是有提升效果的。此篇文章腫瘤本體範圍是由兩位胸腔放射科醫師圈選的，圈選完過了一個月隨機取三十顆再次圈選，並將兩次範圍用 ICC 檢視圈選範圍之特徵相符性，然後腫瘤外的範圍是將醫師圈好之本體範圍使用 AK 軟體讓他取得 3D 膨脹 2mm 後的範圍當作是腫瘤外的範圍，並手動排除



掉血管、氣管等其他組織。取得這些範圍後就是特徵提取的部分，這篇文獻提取的特徵有 histogram, form-factor, GLCM, GLRLM 的特徵，由於訓練集資料的兩類比例有落差，所以會先用 synthetic minority over-sampling (SMOTE) 的方式將訓練集之兩類資料比例變成接近 1:1，通過兩層特徵挑選的方式得到建構模型的特徵，模型建構完成後再將原始的訓練集資料與測試集資料一起放入模型中觀察其結果。而兩層的特徵挑選會先利用 minimum redundancy maximum relevance (mRMR)，找出最小冗餘最大相關的 20 個特徵出來，再用 LASSO 選出最具代表性的特徵並將 lasso 做完得到的每個特徵對應的係數與特徵值相乘後相加，其值稱 radiomic-score，而後將 radiomic-score 放入 logistic regression 得到其 AUC。最後測試集在僅用腫瘤內部特徵與僅用腫瘤外部特徵跟兩者都使用所建立的模型正確率分別為：82.9%、77.1%、88.6%，靈敏度：90%、80%、80%，特異度：80%、76%、92%，及 AUC 為：0.852、0.820、0.876。

最後一篇，由 Guangyao Wu[11]等人所提出之文獻，其資料納入標準為腫瘤需呈現 part solid 的 lung adenocarcinoma，對於 pure GGN 之腫瘤會將其挑掉，不會放入研究當中，另外 solid ratio 也沒有特定的規範。在此篇文獻中僅分出一組訓練集（229 筆）及測試集（68 筆），進行建模與測試。文獻所提出之最佳模型為分別針對腫瘤 solid part 以及 ground glass part 擷取 radiomics 特徵，再挑選具代表性特徵後所建立之模型，並且將其與 clinical model、volumetric model 以及使用不同範圍提取之 radiomics feature 建立的模型進行效果比對。其最佳模型作法為：設定灰階閾值-50HU，將所有腫瘤區域分成兩個部份分別為 Ground glass region (-50HU 以下區域) 以及 solid region (-50HU 以上區域)，透過 RadiomiX Discovery Toolbox，提取兩區域的特徵，提取完後進行特徵選取，作者選擇的特徵選取方式為 ICC、correlation 以及 boruta，經過一連串的挑選後，最後從 solid part 的部分挑選出了三個特徵：(1) 影像於 y,z 軸做 high-pass filter 及 x 軸做 low-pass filter 後，從 NGTDM (neighborhood gray tone difference matrix) 提取的



contrast 特徵，(2) 影像於 x 軸做 high-pass filter 及 y,z 軸做 low-pass filter 後，從 GLDZM (gray-level distance zone matrix) 提取的 DZV (distance-zone variance) 特徵，以及 (3) 影像於 y 軸做 high-pass filter 及 x,z 軸做 low-pass filter 後，從 GLDZM (gray-level distance zone matrix) 提取的 DZV (distance-zone variance) 特徵；而 ground glass region 則挑選出四個特徵：(1) GLRLM 的 SRLGE (short run low gray Level emphasis)，(2) 影像於 x,y,z 軸做 low-pass filter 後，得到位於 10th percentile 的灰階值，(3) compacteness2 以及 (4) 影像於 x,y,z 軸做 low-pass filter 後，計算的 local intensity peak，以上總共七個特徵建立出 invasive 與 non-invasive 的分類模型，得到之結果正確率、靈敏度、特異度與 AUC 分別為：93%、98%、78%、0.9800。

2.2 毛玻璃狀腫瘤之侵犯性分類—深度學習

深度學習透過建立模型的方式讓整理好的資料進入模型訓練，與放射體學最大的不同點為特徵不須先進行提取，而是在模型進行訓練的同時透過 back propagation 的方式不斷為網路參數做更新調整進而產生出適合分類的特徵，所以深度學習比較大的挑戰為建構出一個適合於研究目標用之模型，以優化其不同目的之最後結果與效能。。

2020 年由 Jing Gong [12] 等人提出利用深度學習的模型架構來對毛玻璃狀的肺腺癌進行是否具侵犯性的分類，此篇文獻的資料納入條件與本研究稍有不同，作者並沒有針對 solid ratio 做限制，而是接受任何具有毛玻璃狀的腫瘤影像。文中提到的模型架構主要是基於 residual learning-based CNN network (圖 2.1) 來進行修改，其模型主要有四組卷積池化層後面再接上兩層全連接層，並使用 RMSprop 優化器來減少 cross-entropy 計算出的 loss，而這篇文獻特別的是模型放入的 input 不是拿涵蓋腫瘤本體的 3D VOI 進行訓練，而是將每筆資料從 coronal view, sagittal view 以及 axial view 各切出一張涵蓋腫瘤的 64*64 region of interest，



並將這三張 2D 的 ROI 堆疊形成一張 RGB 影像再放入模型中訓練。另外，為了增加訓練集的數據量，作者利用九種方式來做資料增量：對影像加入高斯雜訊、旋轉 90, 180, 270 度、左右及上下翻轉、位移和 gamma, Logarithmic 矯正，來減少資料過少產生的過擬化 (overfitting) 現象，並且透過資料增量的方式使訓練資料在兩類別比例不要太過懸殊。此篇文獻僅訓練一個模型並用 102 筆 external validation cohort 當作測試集來檢視模型訓練的效果，而其 AUC 可以達到 0.92 ± 0.03 的表現。此外，文中提及實驗的測試數據也同時分別由具有 15 年經驗與 6 年經驗的放射科醫師來進行人為的判斷，判斷依據為 solid portion，佔比越大表示越有機會為 IA，其中 solid ratio 的程度又可以細分成五種 ratings 從小到大的條件為： $\leq 20\%$ (Highly unlikely)、 $20\% \sim 40\%$ (Moderately unlikely)、 $40\% \sim 60\%$ (Indeterminate)、 $60\% \sim 80\%$ (Moderately suspicious)、 $> 80\%$ (Highly suspicious)，最後判定結果 rating 在 1~3 間之資料為 non-IA，而 4、5 為 IA，從數據表現出較資深的放射科醫師做出的判斷與 AI 做出的判斷力是相當的，AUC： 0.92 ± 0.03 ，而另一位放射科醫師的結果為 AUC： 0.90 ± 0.03 。

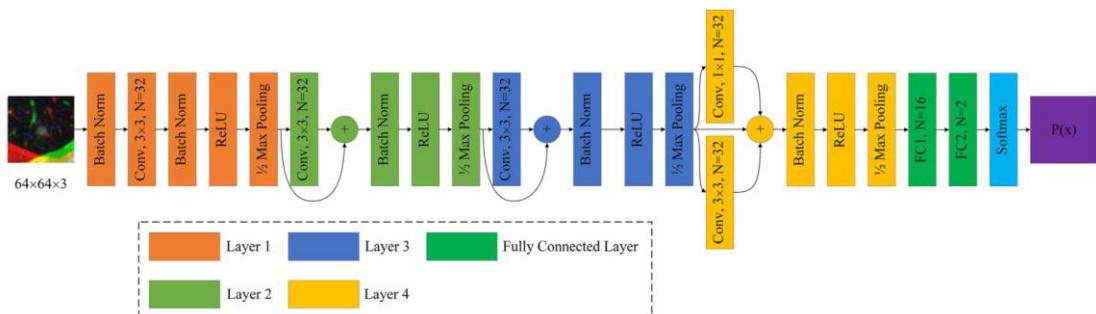


圖 2.1 Residual learning-based CNN network [12]

2.3 毛玻璃狀腫瘤之侵犯性分類—放射體學結合深度學習

2020 年，由 Xianwu Xia[13]等人提出結合深度學習與放射體學的方法來對 non-invasive (MIA/AIS) 與 invasive (IA) 腫瘤分類，文獻中有提到納入的資料為 stage-I 的肺腺癌，其於 CT 影像上呈現 pGGN (pure ground glass nodule) 或 sGGN (part-solid ground glass nodule)，並且最大直徑介於 3mm~3cm 之間。由



於其資料有兩個來源，作者將其中一個來源的資料當作 training 與 validation set，另一個來源的資料當作 testing set，以驗證其提出之方法效能。下圖 2.2 為方法流程圖：

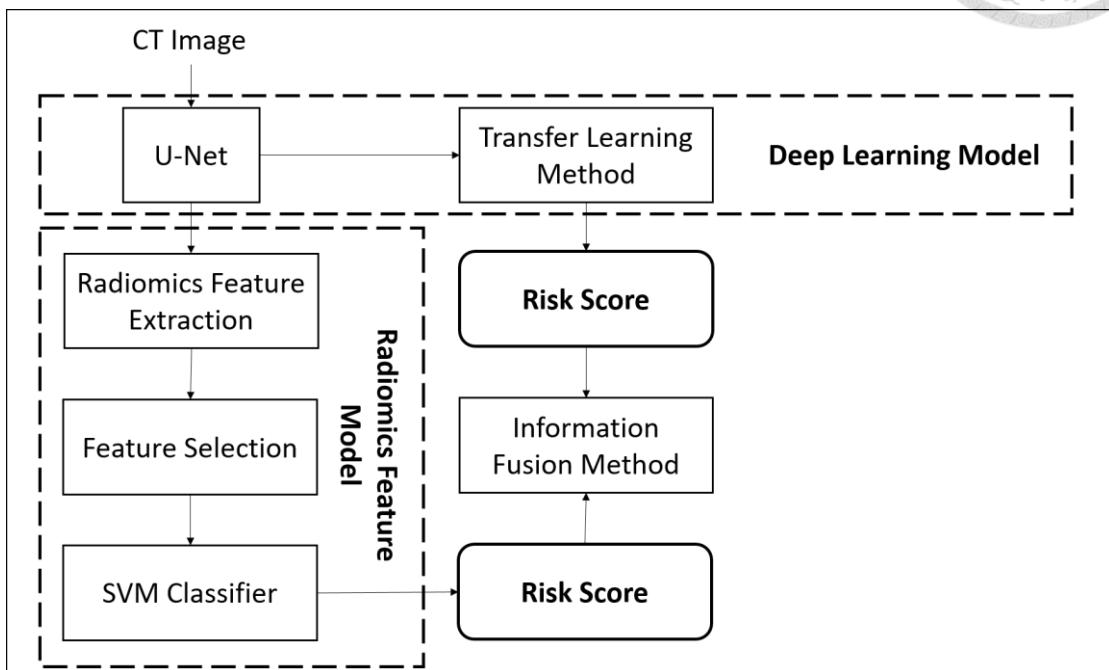


圖 2.2 Flowchart[13]

此文獻提出之方法可以分成三大部分，首先是利用基於 U-net 的 3D recurrent residual convolutional neural network (RRCNN) 拿來做 GGN 的分割，而訓練資料有一部份是來自他們的原始資料，其餘是利用 LIDC-IDRI 公開資料庫裏面的影像並做資料增量來增加訓練的數量。RRCNN 訓練的 loss 是使用 dice coefficient，當 dice 係數越高表示 ground truth 與 precition 重疊度越高，分割效果越好，訓練完 RRCNN 後即成為 pre-train model，接著就進入第二部分，第二部份分成深度學習與放射體學兩個方式同時進行。深度學習的部分作者將 U-net 模型做 feature extraction 的部份的權重進行 transfer learning，保留前面 CNN-pooling，後面再加入兩層全連階層得到分類的模型，接著作者所收集的資料放進模型進行 fine-tune，最後輸出分類結果。另外，放射體學的方法，將利用前面訓練完之 RRCNN model 分割出來的結果當作腫瘤範圍，並對腫瘤內部提取放射體學特徵，其中包含了 laplacian of Gaussian、小波轉換、一階（直方圖）、形狀、紋理（包含



GLCM, GLDM—gray level dependence matrix, GLRLM, GLSZM) 特徵，並利用 ANOVA F-value 的方式做單變量特徵挑選，挑選完 20 個後放入 SVM 進行分類。當 deep learning 與 machine learning 的分類模型都建好了後，就會對兩者的輸出進行融合，其融入方式有三類，第一種是選擇模型分類出來預測之最大值當作最後結果，第二種則相反，取其最小值當作預測機率，第三者則是兩邊預測的結果都會參考，但權重有所不同，而權重從 0.1 至 0.9 都有，若 DL model 結果佔 0.1 則 ML model 就會佔 0.9，兩者權重相加為 1。比較上述之融合方式得到最佳的結果是取 ML model 與 DL model 判斷的預測值中的最大值作為最後的結果，AUC： 0.90 ± 0.03 ，正確率：80.3%，而單純使用放射體學方法跟單純使用深度學習的 AUC 結果為： 0.87 ± 0.04 、 0.83 ± 0.05 ，而其結果也與兩位放射科醫師做比較，其中一位的正確率為 67.7%，另一位為 70.9%。



第三章研究方法與材料

3.1 研究材料

本研究所使用的資料來自台大醫院及新竹醫院，納入的條件包括：(1) 肿瘤在電腦斷層掃描 lung window 下必須是 pure ground glass nodule 或是 part solid ground glass nodule (容許一定比例以下的 solid part) (2) 實質比 (solid ratio)，也就是 CT 影像上較實質的部分最大徑占整顆腫瘤的最大徑經醫師判斷後，必須小於 0.25，因為大多數擁有越多實質部分的腫瘤會越具有侵犯性，因此本研究將實質比較低、較不易判斷的腫瘤挑選出來，針對這些腫瘤去做分類 (3) 肿瘤大小≤ 3 公分，皆為 TNM 分期中歸類為 T=T1 的腫瘤 (4) CT 為細切影像，slice thickness 介於 0.62~1.5mm。

本研究的 CT 影像單一切面大小都是 512*512，KVP 介於 100~140 之間、X-Ray tube current (in μ A) 範圍較廣，在 19 到 647 之間，convolution kernel 也有 18 種不同的方式 (B, B20f, B40f, B40s, B50f, B75f, BONE, CHST, FC08, FC08-H, FC51, 140f\3, 150f\3, L, LUNG, STADARD, YA, YB)，而在量化資料特徵方面統計了兩類別最大直徑與其平均 HU 值之分布，如圖 3.1。

最終挑選完的資料中 Invasive adenocarcinoma 數量是 75 筆，Non-Invasive adenocarcinoma 數量是 175 筆，兩類比為 3:7，表 3.1 為兩類別資料比較表。

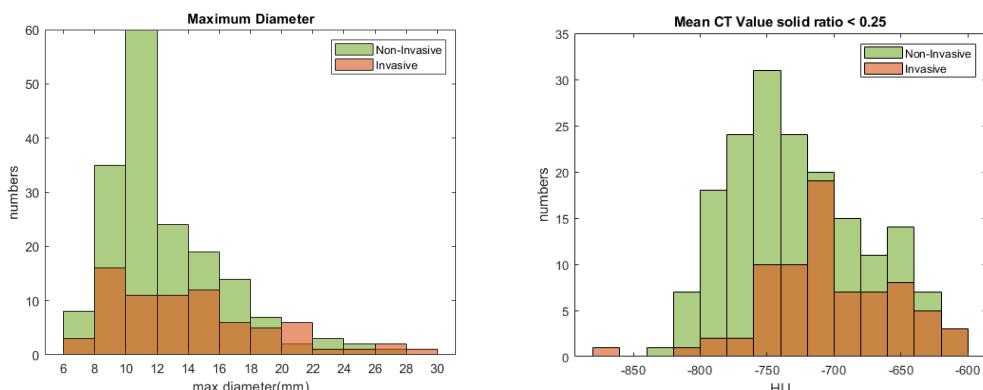




圖 3.1 最大直徑與 mean CT value 分布圖

表 3.1 兩類資料之 CT 參數與最大直徑、mean CT value 比較表

	AIS / MIA(non-IA)	IA
數量	175	75
最大直徑±標準差(mm)	12.6 ± 3.8	14.2 ± 5.1
平均 HU	-726.8 ± 50.1	-706.3 ± 48.3
CT 切面大小	$512 * 512$	$512 * 512$
Convolution kernel	B, B20f, B40f, B40s, B50f, B75f, BONE, CHST, FC08, FC08-H, FC51, 140f\3, 150f\3, L, LUNG, STADARD, YA, YB	B, B40f, B40s, CHST, I40f\3, STANDARD
KVP	100, 110, 120	100, 110, 120, 130
X-Ray tube current	20~640	29~458

3.2 研究方法

本研究採取的方式為結合 AdaBoost 演算法概念的深度學習模型來對毛玻璃狀腫瘤進行是否具有侵犯性的分類，稱之為 AdaBoost Model。此章節會先針對 AdaBoost 演算法作探討，再來對於資料進入模型前所需的前處理做說明，並介紹一篇文獻—AdaBoost-CNN: An adaptive boosting algorithm for convolutional neural networks to classify multi-class imbalanced datasets using transfer learning 探討其如何將 AdaBoost 應用在 deep learning 上，最後再針對本研究所使用的模型架構及如何和 adaboost 結合做更進一步的講解。

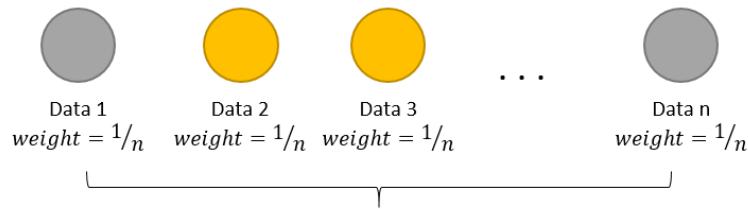


3.2.1. Adaboost 演算法

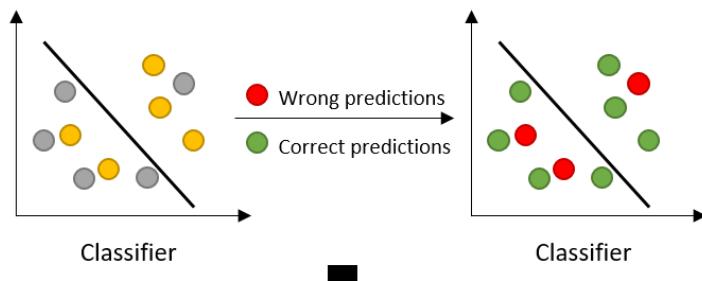
Adaboost 是一個迭代演算法，在迭代的過程中，每筆訓練資料的權重會因前一個分類器對於每筆資料預測的好壞去做改變，以達到需要更著重在分類較差的資料上及對於分類較好的訓練樣本可以給予比較小的懲罰的效果[16]。而除了對權重加以修正外，也有另一種方式是根據修正完的權重去改變訓練樣本集，即對原本訓練樣本重複抽取，因為每筆訓練樣本的權重會經轉換變成抽取資料的機率，所以如果某筆資料的權重比較大（也就是在上一個分類器中分類錯誤的程度較大），那他被抽取到的機率也相對比較大，抽到越多的樣本就會讓下一個分類器對於這筆資料多做學習[17]。每一次迭代所訓練的分類器被稱作是弱分類器，最後會統整這些弱分類器並且依據每個分類器的誤差率給予分類器不同權重（誤差率越高的權重越低），形成強分類器，得到最終的判斷。Adaboost 流程圖如下：



1. Initial weights: $1/n$



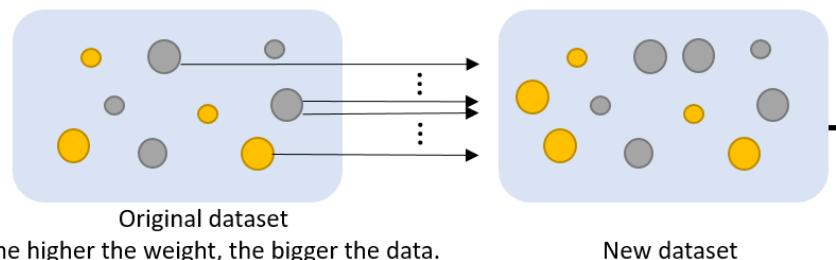
2. Classification



3. Update the data weights

- Data weight ↑
- Data weight ↓

(4. Modified data distribution by new data weights)



5. The strong classifier with k weak classifiers

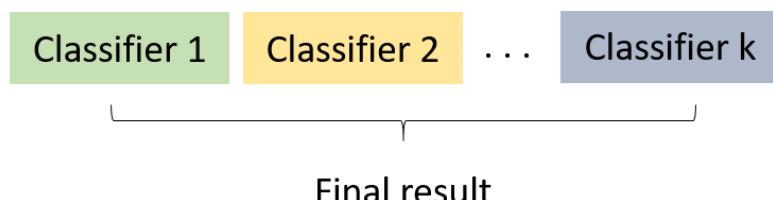


圖 3.2 Adaboost 演算法流程圖

3.2.2. 影像處理

為保持模型輸入資料的一致性及可用性，需統一對影像做正規化，以克服因



影像畫質不一而導致模型訓練效果降低的可能性，另外，將目標影像之原始 DICOM 檔案裁切成較小的 Volume of interest (VOI) 可以避免記憶體不夠之問題，且同時讓模型更能專注在腫瘤周圍資訊，增加訓練效能。由於深度學習是 data driven 的一種方式，所以需要大量的資料進行模型訓練，此時即需要對影像進行增量，讓網絡的參數得以訓練不會快速的過擬化。

此外，因為本研究提出之架構需要使用到腫瘤位置資訊的 binary mask，所以影像分割也是事前所需準備的影像，本研究將透過半自動化的方式將腫瘤區域給分割出來，以對影像所提取出來的資訊做 attention，強化其在腫瘤內部學到的特徵。

3.2.2.1 影像前處理

由於本研究之 DICOM 影像來源不一定相同，而只要來自不同機器所拍攝之 CT 影像就有可能會有不同的拍攝參數，這些參數都會存在 DICOM 標籤 (tag) 內，所以在進行模型訓練前需要利用這些標籤資訊對所有影像做統一化，以避免模型受外在的變因干擾，而限制了其表現效果。

首先，針對像素間距做調整，需提取出 DICOM 中兩個標籤：pixel spacing 與 slice thickness，其分別代表影像在 x、y 方向上的像素間距與 z 軸方向上在掃描時的斷面厚度，透過這兩個標籤可以使影像 resize 成固定的間距，本研究將全部影像都統一 resize 至 $0.6 * 0.6 * 0.6$ 。調整完間距後，會針對影像的像素值做統一，第一步是將拍攝的圖像數值轉換成 CT 值 (HU 值)，此時須提取出 Rescale Slope 與 Rescale Intercept 兩個標籤來做轉換，其轉換公式如式(3.1)：

$$CT = \text{Rescale Slope} * \text{raw image value} + \text{Rescale Intercept} \quad (3.1)$$

第二步為正規化，由於 HU 值之分布廣泛 (-1024~3071)，故在正規化方面會直接將 HU 值除上 4096，並將小於 0 的值取 0、大於 1 的取 1，使得最後影像像素值界在 0~1 之間。最後，再以腫瘤中間的 slice 為中心開出 $64 * 64 * 64$ 的



voxel，作為模型訓練的輸入影像。

3.2.2.2 影像分割

本研究透過實驗室先前所開發之影像分割演算法一步步將影像內的腫瘤區域給分割下來。首先，影像上不同組織（肺實質區域、血管、肺壁...）其灰階差異可以當作是一個分割的依據，於是透過設定的閾值-700HU 當作標準以進行初步的切割，並透過 connected component 六連通的方式取得最大的聯通區域以得到肺壁在影像上的大致位置，與肺實質區域做區分。再者，利用血管分割演算法進行肺實質區域內的血管分割，這個方法是基於 frangi 的方式，透過 frangi filter 檢測圖像內是否具有管狀的結構或纖維，其中 Hessian matrix 是一二階偏微分的對稱矩陣，用於影像上主要是描述特定位置上的函數值與其相鄰位置之函數值變化，而管狀結構裡面的三個 eigen value(λ_1 、 λ_2 、 λ_3)具有以下特性： $\lambda_1 \approx 0$ ， $|\lambda_2| \gg |\lambda_1|$ ， $|\lambda_3| \approx |\lambda_2|$ ，以判斷血管（管狀物）。接著，將針對腫瘤部分進行 level set 演算法，其原理為透過每個點所在位置的梯度決定其行進的方向與速度，當梯度較小表示附近區域的灰階大小雷同，可以視為還存在於腫瘤範圍內，進而加快移動速度；反之梯度越大，表示此點所在的位置接近於邊緣，移動的幅度就會減少，透過手動調整方式依照腫瘤大小去調整迭代次數，讓曲線上的點在迭代的過程中不斷移動位置，直至迭代結束後得到最終曲線範圍。為避免 levelset 圈選之腫瘤範圍與肺壁或血管重疊，所以須扣除掉血管與肺壁的位置，以得到腫瘤分割結果，最後再針對分不好的部分進行微調。圈選完腫瘤範圍後，由醫師進行最後確認，以得到最終分割結果。下圖為影像分割前後圖（圖 3.3、圖 3.4）：

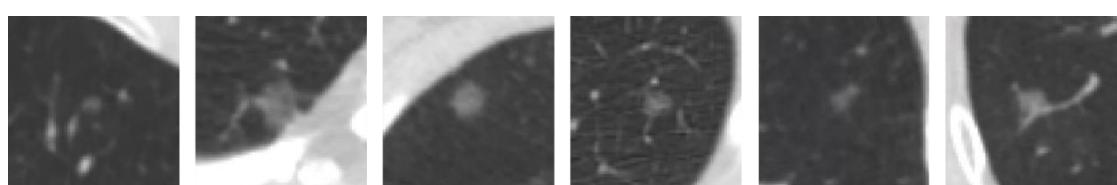


圖 3.3 原始影像圖

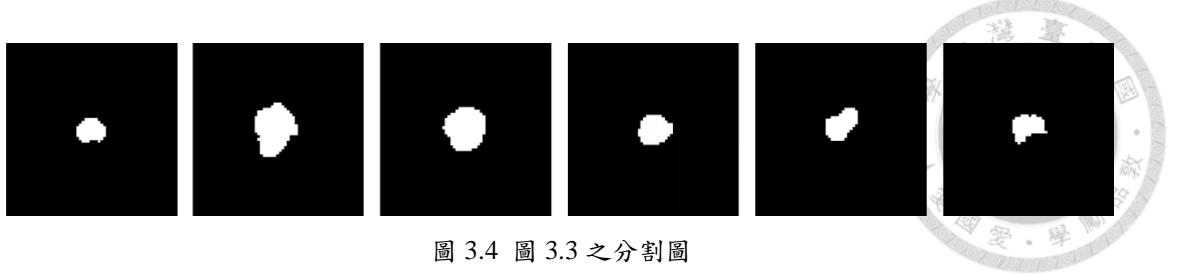
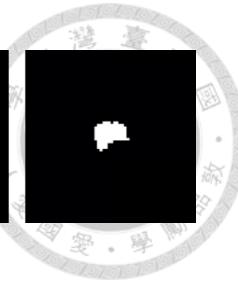


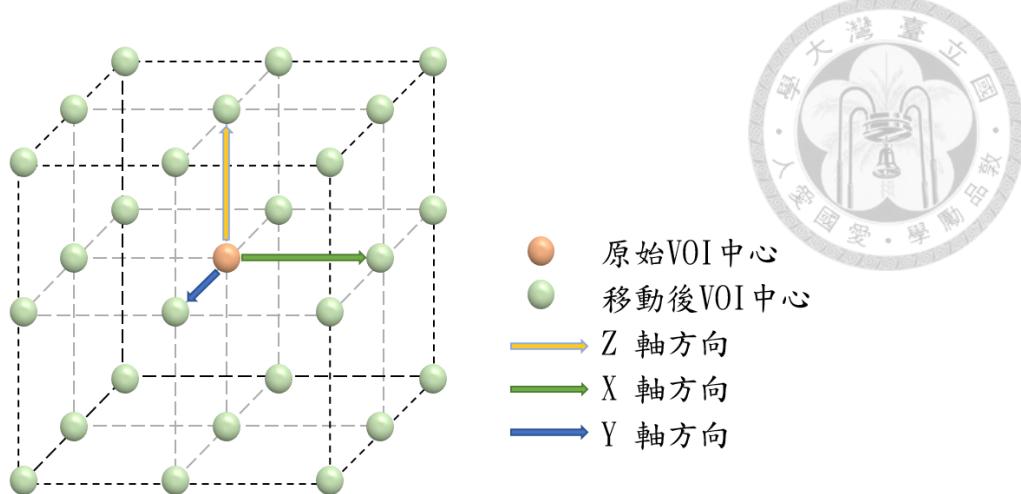
圖 3.4 圖 3.3 之分割圖



3.2.2.3 資料增量

Deep learning 為一種需要大量訓練資料來進行建模的方法，由於訓練模型的過程中要學習大量的參數來獲得最佳的網路參數，所以放入較多數量的資料可以得到更多資訊讓模型可以學習，減少產生 over-fitting 的現象。在醫學影像上的取得數量有限，因此需透過資料增量的方式將手上有的原始影像去做處理以創造更多的影像，來彌補資料不足的問題。

本研究透過位移（Shift）與旋轉（Rotation）來產生新的資料，在位移方面有 x、y、z 三軸可以做移動，而本研究在 x 與 y 兩軸設定可移動之步長為 4 pixels 與 8 pixels，相對應 z 軸可移動之步長為 3 slices 與 5 slices，假設現在 x、y 可移動 4 pixels，則 z 軸可移動 3 slices，總共可移動之排列組合為 $(x, y, z) : (4, 0, 0), (-4, 0, 0), (0, 4, 0), (0, -4, 0), (0, 0, 3), (0, 0, -3), (4, 4, 0), (-4, -4, 0), (4, -4, 0), (-4, 4, 0), (4, 4, 3), (-4, 4, 3), (4, -4, 3), (4, 4, -3), (-4, -4, 3), (4, -4, -3), (-4, 4, -3), (-4, -4, -3), (4, 0, 3), (-4, 0, 3), (4, 0, -3), (-4, 0, -3), (0, 4, 3), (0, -4, 3), (0, 4, -3), (0, -4, -3)$ 共 26 種，如下圖所示：



總結以上移動與旋轉方式，可移動之組合有兩種：(1) $x, y : 4 \text{ pixels}$, $z : 3 \text{ slices}$ (2) $x, y : 8 \text{ pixels}$, $z : 5 \text{ slices}$ ，並且每移動完一次都有四種方向可以進行旋轉，分別為： 90° 、 180° 、 270° 、 360° ，所以每筆資料在做完增量資料的新增樣本數目共有 $2(\text{兩種移動組合}) * 26(\text{位移的 26 種方向}) * 4(\text{旋轉的四個角度}) = 208$ 種。

由於原始兩類別資料比為 $3:7$ ，所以當原始資料分好5個fold後，每個fold的訓練集資料都會加入這些原始資料所產生之增量資料當中抽取比例為 $3:7$ 的樣本放入訓練集一起給模型做訓練。

3.2.3. 模型架構

深度學習建立之分類模型大致可以拆解為兩個階段，第一階段主要功能就是對影像中的資訊做特徵提取（Convolution layer, max pooling layer），越接近輸入端的layer提取出來的特徵多為scale較小的特徵、屬於小的細節（low-level features），像是邊緣、角、梯度等等，而越到後端（越深層），模型可以根據前面學到的 low-level feature 上，再更深入學習進而得到更具有語義、更複雜、更抽象的特徵資訊（high-level features）；模型第二階段的功用，是將第一階段提取之特徵當做依據，進行目標物的分類（fully connected layer）。

此章節將先針對提出的模型架構進行拆解，說明基礎架構的功能與做法，其



中基礎架構分別有 Depth-wise Convolution (Convolution layer)、max/ global average pooling layer、fully connected layer、activation function、squeeze and excitation layer、attention layer，了解各子架構的功能後再介紹本研究所提出之整體模型架構—AdaBoost Model。

3.2.3.1 Depth-wise Convolution Layer

此節將針對基本的 Convolution layer 做原理說明後，再對本研究之模型子架構中的 depth-wise convolution 作介紹。Convolution layer 是由兩個步驟所組成：滑動及內積，即將輸入至此層的影像或是 feature map 與在上面做滑動的 kernel（或稱 filter）做內積，以特徵濾波器的概念對 feature map 做特徵的提取。

Convolution layer 之數學式如下式(3.2)：

$$x_j^l = f \left(\sum_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_j^l \right) \quad (3.2)$$

M_j	第 M_j 個 feature map
x_i^l	第 l 層，feature map 上滑動到第 i 塊與 kernel 內積
x_j^l	x_j^{l-1} 做完卷積的結果
k_{ij}^l	卷基層的 kernel
b_j^l	bias
$f(\cdot)$	激活函數

進行卷積之時所需要的參數主要有：kernel number 與 kernel size。kernel number 調整對影像或 feature map 做卷積的 kernel 數量，其數量與做完卷積運算後產生出來的 feature map 數量相同；另一個參數 kernel size，為 kernel 本身的大小，尺寸越大，模型所需要學習的參數也越多。卷積運算做法如圖 3.5：

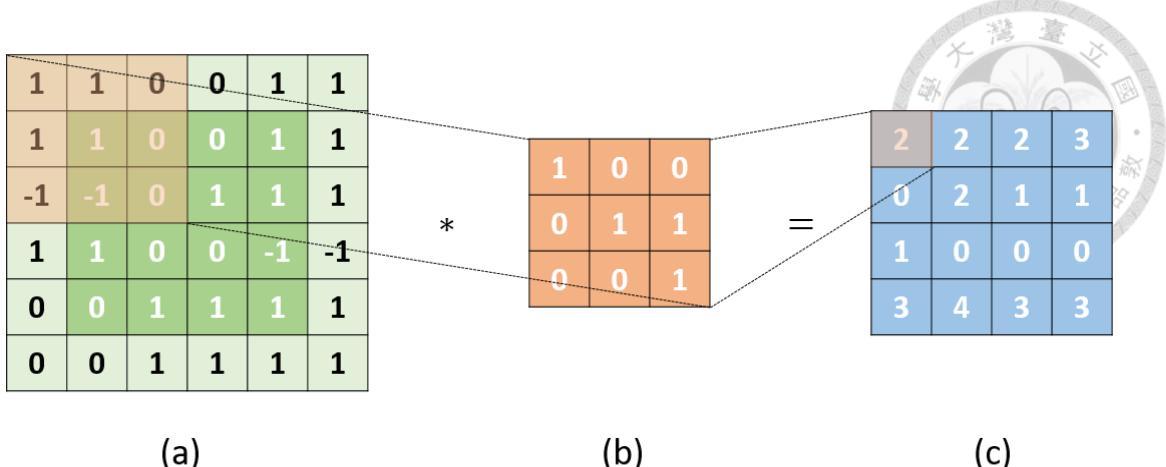


圖 3.5 Convolution 卷積運算 (a)是原始影像 4×4 ，(b)為 kernel，(c)為原始影像做完卷積之結果

圖 3.5 (a)是 4×4 的 feature map (x_j^{l-1})，為保持做完卷積後仍維持相同的 feature map，進行卷積前須對影像做 padding (即淺色外圍，以黑色數字表示)，而 padding 的數字是用鄰近的像素質來進行填補。圖上有橘底的部分是 x_i^l ，將與 kernel 進行卷積。(b) kernel (k_{ij}^l)，(c) feature map 做完卷積的結果 x_j^l 。

Depth-wise convolution[18] (又稱作 separate convolution) 的目的是在希望不影響輸出結構的狀況下減少運算量，可以分成兩步驟：depth-wise convolution 與 point-wise convolution。做法與一般卷積相異的地方在於它將 feature map 的 channel 進行拆解，拆解後每個 channel 的 feature map 就會各自和任意大小的 kernel 進行卷積（此為第一個步驟：Depthwise convolution），再 concatenate 被拆解的 feature map 後進行一次 $1 \times 1 \times 1$ 的卷積運算。depth-wise convolution 最大的優點是可以增加模型的運算效率、減少運算量。其運作範例如圖 3.6：

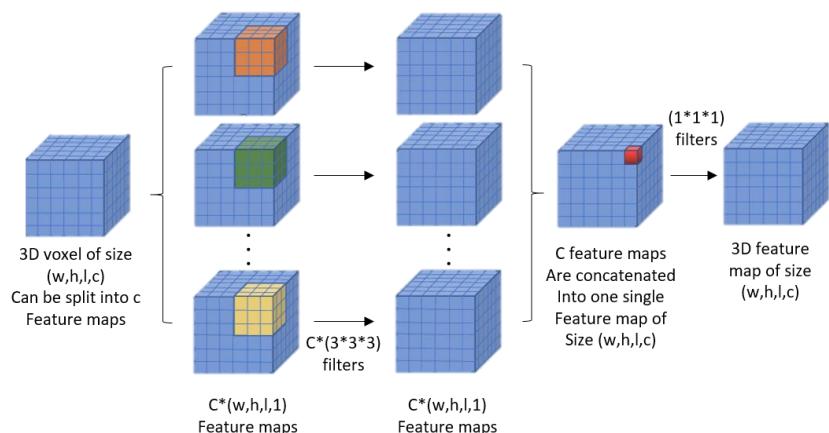


圖 3.6 Depth-wise Convolution[18]

於本研究 depth-wise convolution 後面都會在接上 batch normalize 與 activation function，以確保特徵不會太分散，使的特徵所給予的幫助不會有不平衡的情況。

3.2.3.2 Max /Global Average Pooling Layer

池化層的功能即將輸入的 feature map 的尺寸縮小，以減少維度並保留重要的特徵。池化層的優點為不僅不需要學習任何參數還能使輸出的 feature map 尺寸縮小以減少後面模型學習的參數量，同時保留一樣具代表性的特徵。於本研究中，共使用兩種不同的池化層，分別為最大化池化層與全局池化層，兩者在實作方面也有所不同。

首先，最大池化層運作方式與卷積層一樣有 kernel 在 feature map 上做滑動，但其滑動範圍並不會互相重疊（通常 kernel 大小為 2×2 ，步數為 2，反觀卷積層之步數多為 1），並且每滑動一次便提取 kernel 涵蓋範圍之最大值，示意圖如圖 3.7：

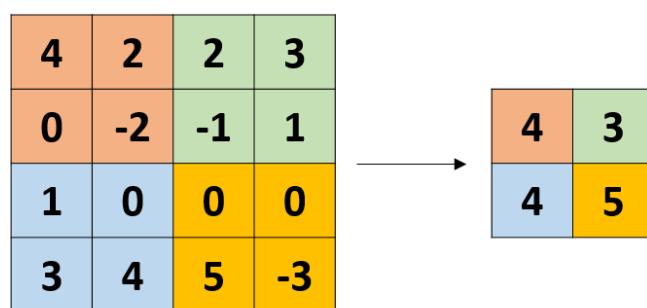


圖 3.7 Maxpooling 示意圖

全局池化層[19]，與最大化池化層相同的地方在於一樣有 kernel 會在影像上做滑動，但 kernel 的大小為每個 channel 其 feature map 的大小，也就是 kernel 每做一次池化就會將一個 channel 裡所有的參數變為一個數值，而此數值是由全部數值的平均求得的，所以若輸入端的大小為 $w \times h \times \text{channel}$ ，則做完全局池化後的大小即為 $1 \times 1 \times \text{channel}$ 。全局池化層相較於 flatten 可以減少全連階層訓練的參數量，以減少過擬化的情況。如所示圖 3.8：

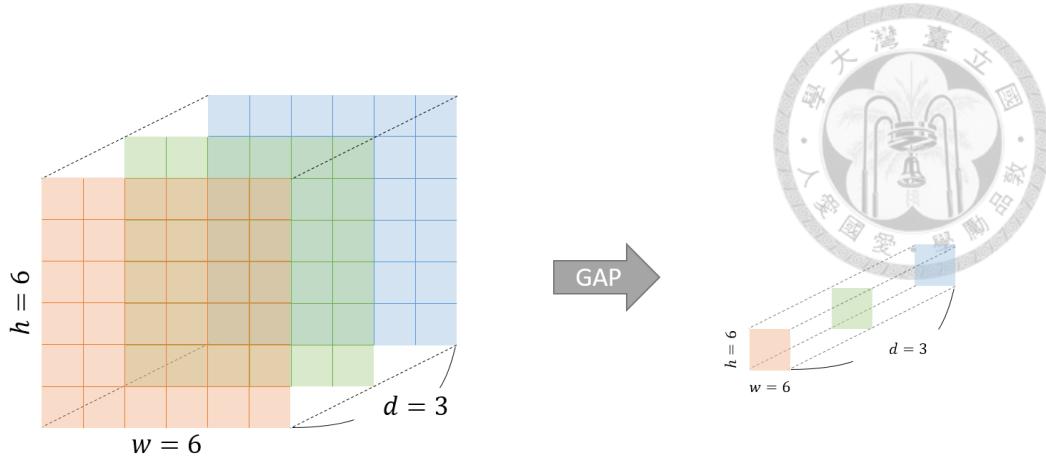


圖 3.8 Global Average Pooling

參考 <https://www.cnblogs.com/hutao722/p/10008581.html> 圖繪製

3.2.3.3 Fully Connected Layer

經過數層卷積和池化層後，通過 flatten（或是 global average pooling）會使原本高維的 feature map 展開成大小為 $(w * h * l * c)$ 的一維特徵向量（做完 global average pooling 後大小則為 c ），每一個節點都包含了模型前半部所學習到的語義和空間資訊，接著這些特徵即可進入 fully connected layer（全連接層），其神經元會與前一層的所有神經元相互連接，在神經元中做組合，為常規類神經網絡 [20]，並於各層全連接層中慢慢萃取出有效特徵以對映到樣本的真實類別來達到分類效果，其主要做法就是透過矩陣向量的乘積。這步驟主要調整的參數為節點數量，多數做法會從數量較大的節點慢慢傳遞到數量小的節點，最後的節點數則多為目標分類的類別數量並通過後面將進行介紹的 activation function（激活函數）達到分類器的效果。其運算式(3.3)與圖 3.9 如下所示

$$x_j(x) = f\left(\sum_{k=1}^n W_{jk}x_k + b_j\right) \quad (3.3)$$

W_{jk}	為連接第 j 、 k 層之權重，且 $k = j - 1$
x_k	為第 k 層之 neuron
b_j	為 j 層之 bias

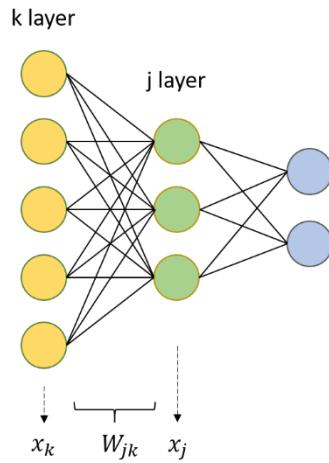


圖 3.9 Fully connected layer

3.2.3.4 Batch Normalize、Activation function

Batch normalize 的做法是對每個 batch 的 output feature 取得全部的平均值（式(3.4)）與變異數（式(3.5)）後，將各個 feature map 減去平均並除上標準差，使特徵呈常態分佈，以落在一定的空間範圍內減少分散程度，式(3.6)。

$$\mu = \frac{1}{n} \sum_{i=1}^n z^i \quad (3.4)$$

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (z^i - \mu)^2} \quad (3.5)$$

$$\tilde{z} = \frac{z^i - \mu}{\sigma} \quad (3.6)$$

n	一個 batch 裡面的資料數量
z^i	第 <i>i</i> 筆資料

Activation function 有很多選擇，像是最常使用的 relu、sigmoid、leaky relu...，本研究接續在卷積層與 batch normalize 後的激活函數是選擇 relu[21]，圖形如圖 3.10，只保留正值的部分，式(3.7)。relu 比起 sigmoid 在做 backpropagation 的時候可以避免梯度消失的問題並且運算也較為簡單（僅需判斷輸入質是否為正）。在本研究 fully connected layer 最後要對映至目標函數類別的 output 前也會通過一個激活函數—softmax，如式(3.8)，可以針對多類別去給予每一個目標機率值，最後即挑選機率值最高的類別當作模型預測的目標類別。

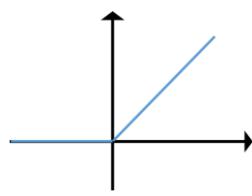


圖 3.10 relu



$$f(x) = \max(x, 0) \quad (3.7)$$

$$f(z_j) = \frac{z_j}{\sum_{k=1}^K e^{z_k}} \quad (3.8)$$

Batch normalize 與 activation function 通常會在卷積層後面出現，其目的是先將每一個 batch 的資料做正規化，使其呈平均值為 0、變異數為 1 的常態分佈，接著進入 activation function 對進入的 feature map 做激活，為神經網絡增加非線性的關係以避免所有模型的輸入輸出無法脫離線性關係，對於深度學習就多了一層意義，如圖 3.11。

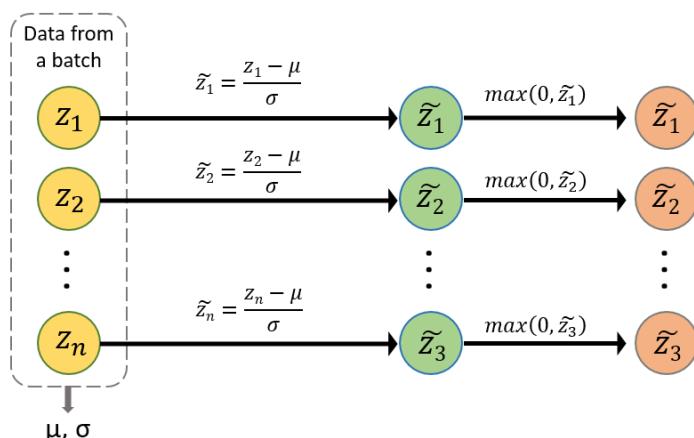


圖 3.11 Batch Normalize 與 Activation layer

3.2.3.5 Squeeze and Excitation Net

Squeeze and Excitation Net (簡稱 SENet) [22]主要的目的為學習不同 channel 各自的重要性，以提高模型的表現性，而其做法就如名稱包含兩個步驟：squeeze 以及 excitation 的部分。

首先，輸入的 feature map 利用 global average pooling 將每一個 channel 的 feature map 取平均變成一個值，使得原本的大小 $w * h * \text{channel}$ 轉為 $1 * 1 * \text{channel}$ 。



channel的特徵向量，此為第一個步驟：squeeze 壓縮，此特徵向量接著進入全連接層與非線性激活函數來學習每一個 channel 各自的重要程度，即為 excitation，提取完每個 channel 的重要性後將此 $1 * 1 * c$ 的特徵向量放大回原始 $w * h * c$ channel大小，以便與一開始進入的 feature map 做相乘，來給予每個 channel 於 SENet 中所學習到的各自權重，以表示每一層的特徵重要性。SEnet 如下圖 3.12 所示：

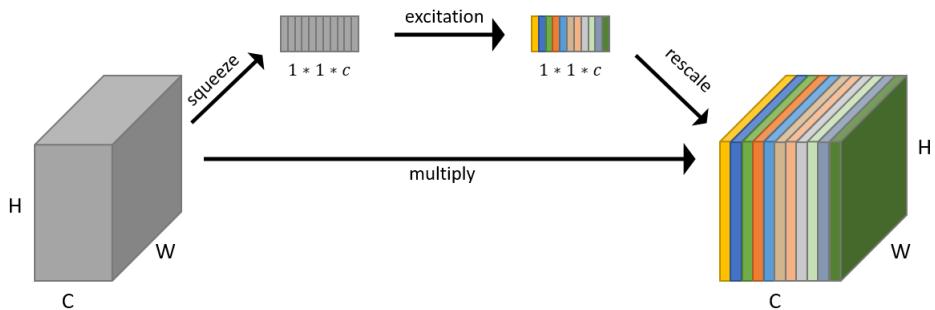


圖 3.12 SE net[22]

3.2.3.6 Attention Layer

本研究有一個重要的子架構為 attention layer，利用影像處理時產生出的分割結果（binary mask）來加強模型學習。由於本研究目標影像屬於 solid ratio < 0.25 之毛玻璃狀腫瘤，在電腦斷層掃描影像上較為黯淡不清楚，如圖 3.13 所示，所以需要另外的方式輔助模型訓練。其做法為對原始訓練模型的 feature map 乘上 binary mask 以針對腫瘤內部區域做加強學習，並去除腫瘤外的雜訊（血管、肺壁等等...），著重於腫瘤本身的特徵給後面的神經網絡進行更 high-level 的特徵提取。

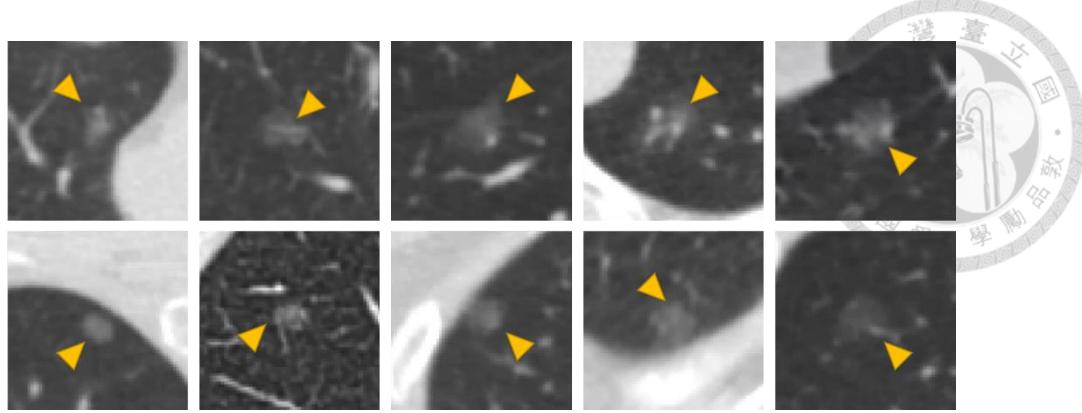


圖 3.13 毛玻璃狀腫瘤 CT 影像

於本研究提倡的模型當中，attention layer 共使用了兩次，分別加在模型最前面兩個 max pooling layer 後面，如下圖 3.14：

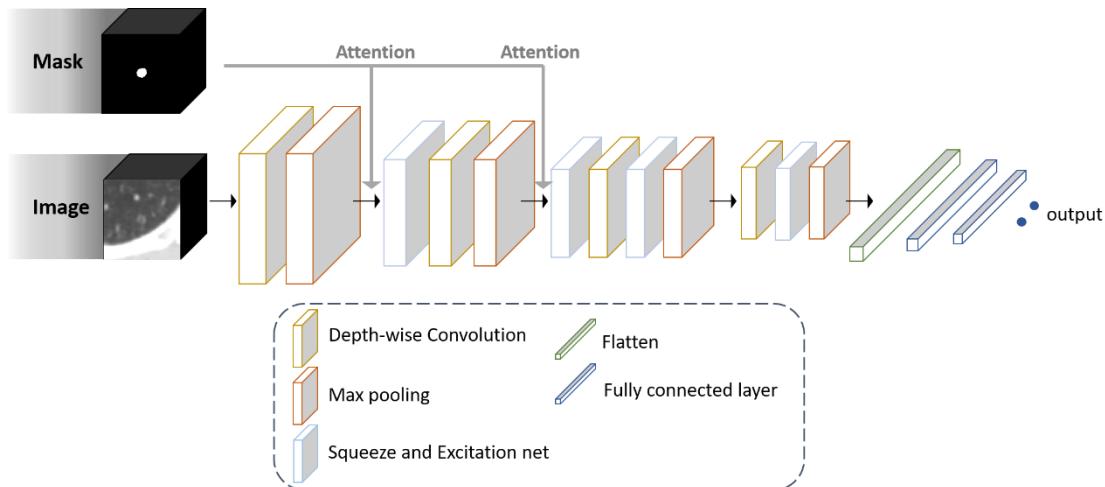


圖 3.14 AdaBoost model

3.2.3.7 Proposed Model

根據 3.2.3.1~3.2.3.6 中所有小節介紹之基本模型架構拼湊，可以得到本研究之整體模型架構，更為詳細之模型架構圖如圖 3.15，以 max pooling layer 作為區隔，主要分為五個 block。

首先 input 為一個切割好的 $64 \times 64 \times 64$ 的 VOI 影像，進入模型後會通過 filter number 為 4 的 depth-wise convolution，並於通過 max pooling 將影像尺寸縮小後與 binary mask 做 attention，將前面所學習到腫瘤以外的雜訊給排除，使模型著重學習腫瘤內部資訊。而後進入 Block 2 的 SENet，將每個 channel 各自學習到的重



要程度與原本的 feature map 做相乘，輸出的 feature map 同樣再通過 filter number 為 4 的 depth-wise convolution、max pooling 與 attention。Block 3 與 Block 2 相異的地方在於 depth-wise convolution 與 max pooling 之間有多加入一個 SENet；再來進入的 Block 4 一樣由 filter number 為 4 的 depth-wise convolution、SEnet、max pooling 所組成，前面四個 Block 即為特徵提取的部分，接著會進入最後分類器的階段。透過 Flatten 先將原本的 feature map 拉長成一維的特徵向量，並進入 fully connected layers 做最後的分類，通過 softmax 後 output 出來的機率值即可取最大值當作模型最後預測的類別。此研究所提出之模型架構之所以將 filter number 都設為一個比較小的數目 4 是為了減少模型的參數，由於本研究所蒐集之資料為數不多，過多的訓練參數會導致過擬化的情形相當嚴重，使得測試集的資料測試結果不佳，故選擇較小的 filter number。整體模型架構總共有 9476 個訓練參數。

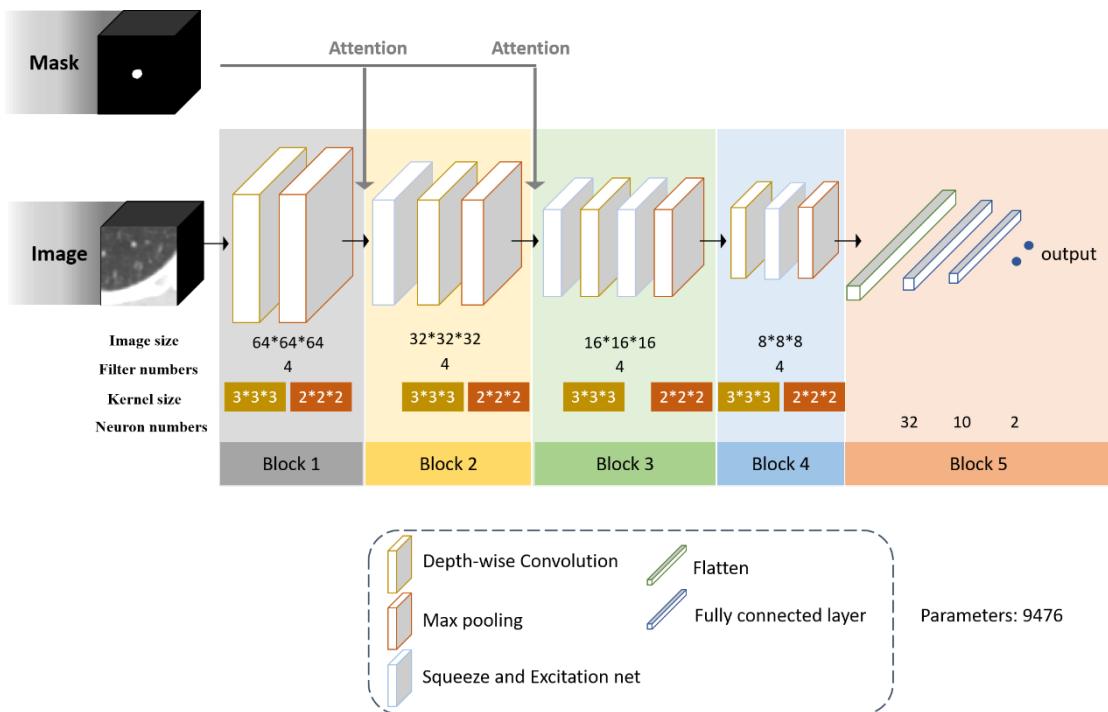


圖 3.15 Adaboost Model

3.2.4. 深度模型更新資料與權重

將 AdaBoost 應用在深度學習這部分是參考 Taherkhani, 2020 [23]這篇文獻來



加以改善的。Adaboost 這個演算法其實就是 ensemble 的方式，在迭代的過程中不斷改變訓練樣本的權重然後利用這些調整過的樣本再繼續訓練模型。迭代結束後，再將測試樣本放入前面 M 個訓練完的模型中去做預測，最終結果也是由這 M 個預測結果去做統整。對於深度學習的 AdaBoost CNN 比起傳統的 AdaBoost 來說，由於能對資料進行資料增量所以相較下比較不易受資料不平衡的問題影響，並能接受大量的訓練資料。

作者的做法是在第一次的模型 ($C^{m=1}$) 訓練中，給予每一筆訓練資料 (x_1, \dots, x_n) ；其對應類別是： c_1, \dots, c_n ； n 為訓練樣本資料數量；共有 K 類別) 一樣的權重 (d_1, \dots, d_n) ，也就是 $1/n$ ，模型的初始參數也是隨機給定的，設定完這些參數後就進行模型訓練，訓練完最後會得到一個輸出結果 K 維向量，是針對每一個類別的預測機率 $(P(x_i) = [p_k(x_i)], k = 1, \dots, K)$ ，所以具有最高機率值的類別即為這個模型預測出來的最終類別。在進入第二次模型訓練前，會先透過前一次模型分類出來的結果去更新第二次訓練時所有訓練樣本的權重，其更改權重的公式如式(3.9)：

$$d_i^{m+1} = d_i^m \exp\left(-\alpha \frac{K-1}{K} \gamma_i^T \log(P^m(x_i))\right), i = 1, \dots, n \quad (3.9)$$

α	learning rate
γ_i	第 i 個資料真實類別向量
$P^m(x_i)$	第 i 個資料在第 m 個(前一個)模型預測結果向量

此公式是參考 SAMME.R 演算法[24]，利用預測的機率來更新每筆資料的權重，在 exponential 裡面有一部份是由兩個向量相乘： $-\gamma_i^T \log(P^m(x_i))$ ，前者 γ_i 為第 i 個資料真實類別向量，後者為第 i 個資料在前一個模型預測結果向量，當某類別預測的機率較高時，其機率值取完 log 後的數值再加一個負號會是比較小的正值，反之預測機率低時，其機率值取完 log 的數值再加上一個負號就會是比較大的正值，所以當預測機率與真實類別相符的情況下， $-\gamma_i^T \log(P^m(x_i))$ 的內積結果加負號會比較小，將其取過 exponential 之後再乘上原始的權重值就會減少原本的權重；若不相符 $-\gamma_i^T \log(P^m(x_i))$ 就會得到較大數值，進入 exponential 再乘上原始



權重便會得到相對較大的權重，以便在新的模型訓練當中多著重在這些預測與實際類別不相符的資料。對訓練資料做完權重更新後會將新的權重做正規化，使的權重總和為 1，即為最後更新的權重，並丟入下一個模型當中。在這邊，作者提出了 transfer learning 的方法，並提到比起在新的迭代中完全重新訓練一個模型，將前一個訓練模型訓練出來最佳的參數 transfer 至新的模型有較好的效果，讓新的模型同時保留上一個模型所學到的資訊又可以透過更改過後的訓練資料權重對上一層分類器分不好的地方多做處理來微調網路 weight。在深度學習的簡化流程圖如圖 3.16：

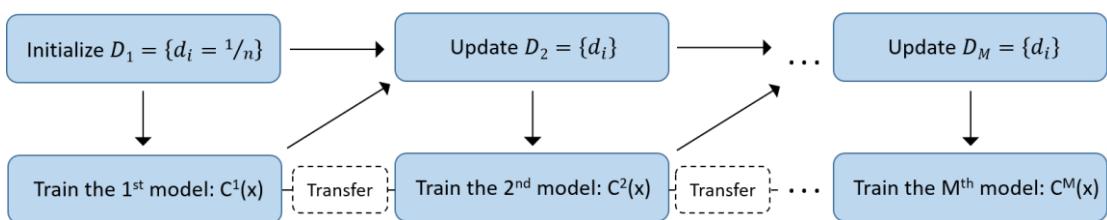


圖 3.16 AdaBoost CNN Flowchart[23]

由於 Adaboost 是 ensemble 的方法，所以作者提出計算每個模型預測結果的方式，其透過式(3.10)計算：

$$C(x) = \operatorname{argmax}_k \sum_{m=1}^M h_k^m(x) \quad (3.10)$$

其中 $h_k^m(x)$ 為式(3.11)：

$$h_k^m(x) = (K - 1) \left(\log(p_k^m(x)) - \frac{1}{k} \sum_{k=1}^K \log(p_k^m(x)) \right) \quad (3.11)$$

x		input
$P^m(x)$		k^{th} element of the output vector of the m^{th} model

最後結合各模型結果即可得到最終預測目標。

本研究沿用這篇文獻中更新權重的方式以及 transfer learning 的部分，利用式(3.9)將分類結果較好的資料權重調小、比較不好的資料其權重加大以便下一個模型訓練時可以針對他去做學習。另外 transfer learning 的方式讓每次訓練新的模型之前，都可以先將前一次的訓練參數 transfer 至新的模型，讓模型可以保留在前



一個模型所學到的特徵並透過訓練前更新的權重以著重這些分不好的資料，多去做學習讓網路參數做微調。除了參考文獻的更改權重外，本研究透過更新完的權重向原始訓練資料做抽取，讓比較難以辨別的樣本資料有機會多增加一些數目，達到讓模型多去學習那些不好學習的樣本的目的。下面將對抽取資料的方式進行介紹：透過前面式(3.9)調整好每個訓練樣本的權重過後，會透過標準化讓所有樣本權重調整結果總合為 1，而這個權重可以被拿來當作是抽取每個樣本的機率，舉例來說，如果現在的樣本資料有五個，權重分別是 0.1, 0.2, 0.3, 0.35, 0.05，這五個權重值即可產生出五個區間： $[0, 0.1)$, $[0.1, 0.3)$, $[0.3, 0.6)$, $[0.6, 0.95)$, $[0.95, 1]$ ，此時就可以從 0~1 之間隨機取一個數，再看這個隨機值落在哪一個區間，（此處可以發現當樣本資料的權重越大，區間的大小也會越大，則被抽中的機率也會提升）即抽取在此區間的那筆資料丟入訓練樣本中，然後重複抽取 1500 次，所以每次的訓練除了原本的訓練樣本之外還會再增加 1500 筆資料，去給新的模型做訓練，每一次迭代都會根據新的樣本權重產生區間並抽取新的 1500 筆資料。抽取樣本的簡易流程如下圖 3.17：

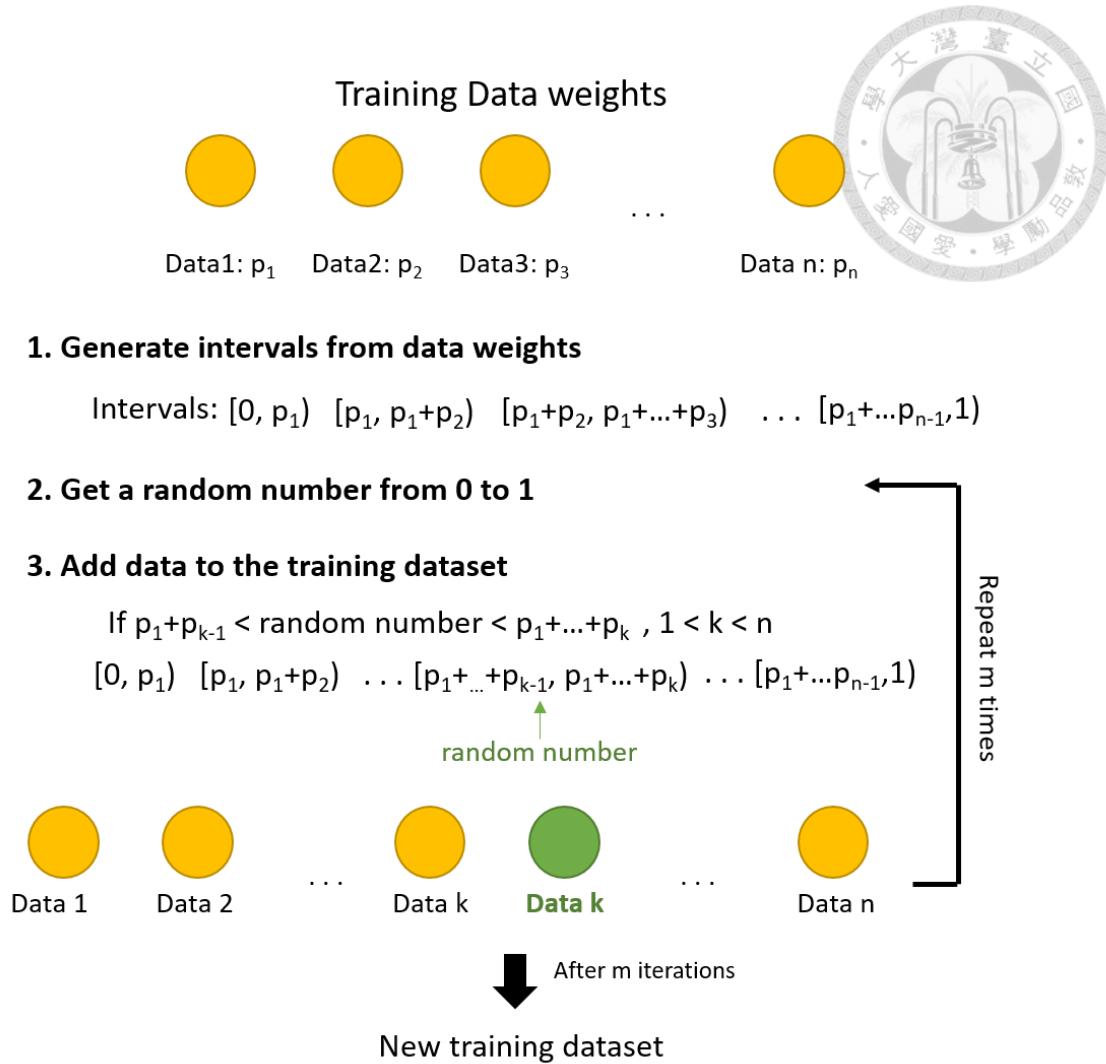


圖 3.17 AdaBoost Model 抽取樣本流程圖

3.3 效能評估

3.3.1 效能指標

為評估模型分類的預測結果好壞，通常使用的測量指標有（1）正確率（Accuracy）、（2）靈敏度（Sensitivity）、（3）特異度（Specificity），而這三項指標之計算是透過分類結果區分的四種可能情況所計算的，分別為真陽性（True positive, TP）、偽陽性（False positive, FP）、真陰性（True negative, TN）、偽陰性（False negative, FN），這四個情況可以以混淆矩陣（圖 3.18）來表示：



Confusion matrix		Actual	
		Positive	Negative
Predict	Positive	True positive	False positive
	Negative	False negative	True negative

圖 3.18 混淆矩陣

本研究對應之真陽性為實際與預測都為 Invasive Adenocarcinoma

、真陰性為實際與預測是 Non-Invasive Adenocarcinoma (minimally invasive adenocarcinoma/adenocarcinoma on situ)、偽陰性為實際為 Invasive Adenocarcinoma 但預測出來的是 Non-Invasive Adenocarcinoma (minimally invasive adenocarcinoma/adenocarcinoma on situ)，及偽陽性實際為 Non-Invasive Adenocarcinoma (minimally invasive adenocarcinoma/adenocarcinoma on situ) 但預測結果是 Invasive Adenocarcinoma。利用這四種情況即可獲取前面提到的三項指標：

(1) 正確率 (Accuracy)：可正確預測兩個類別的能力，即實際與預測相同（都為 Invasive Adenocarcinoma 或是都為 Non-Invasive Adenocarcinoma）佔所有預測結果的比例，可表示為式(3.12)：

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (3.12)$$

(2) 睿敏度 (Sensitivity)：表示在實際結果為 Invasive Adenocarcinoma 的樣本當中，可以正確被預測的比例，式(3.13)：

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (3.13)$$

(3) 特異度 (Specificity)：表示在實際結果為 Non-Invasive Adenocarcinoma 的樣本當中，可以正確被預測的比例，式(3.14)：

$$\text{Sensitivity} = \frac{TN}{TN+FP} \quad (3.14)$$



另外，還有一項重要的評估指標可以更全能的評估檢測模型的分類鑑別力：Area Under Curve(AUC)，即為計算曲線（Receiver operator characteristic curve, ROC curve）下的面積占總面積的比例，數值介於 0~1 之間。而 ROC curve 的定義是在各種決策門檻（discrimination threshold）下，比較真陽率（Sensitivity or recall, Y 軸）與偽陽率（1-Specificity, X 軸）之間的變化，每個門檻都能構成一個點，連接起來便能形成曲線，曲線上越靠近左上角的點（0, 1）可以得到錯誤歸類最少的切點，因為此切點的真陽率是最大的且偽陽率是最小的。AUC 為 1 時，表示模型可以進行完美預測；當介於 0.5~1 之間，表示模型相較於隨機猜測而言，是更具有判斷能力；對於 AUC 介在 0~0.5，則表示模型預測結果相對於隨機預測是較不具有參考價值。

此外，本研究為避免過度依賴某些特定的訓練或測試資料而導致訓練效能的偏差，採用的是 5-fold cross validation，也就是將資料隨機拆成 5 個集合來做交叉驗證，每個回合都由其中三份當作訓練集樣本、一份為驗證集樣本、最後一份為測試集樣本，三者的數量比例為 3:1:1，其中驗證集樣本可視為從訓練即樣本抽取出來的一部份資料，由於每一個集合都需當過測試集樣本，所以訓練五個回合後就可以對五組測試集樣本的結果做出統整。如下圖 3.19 所示：

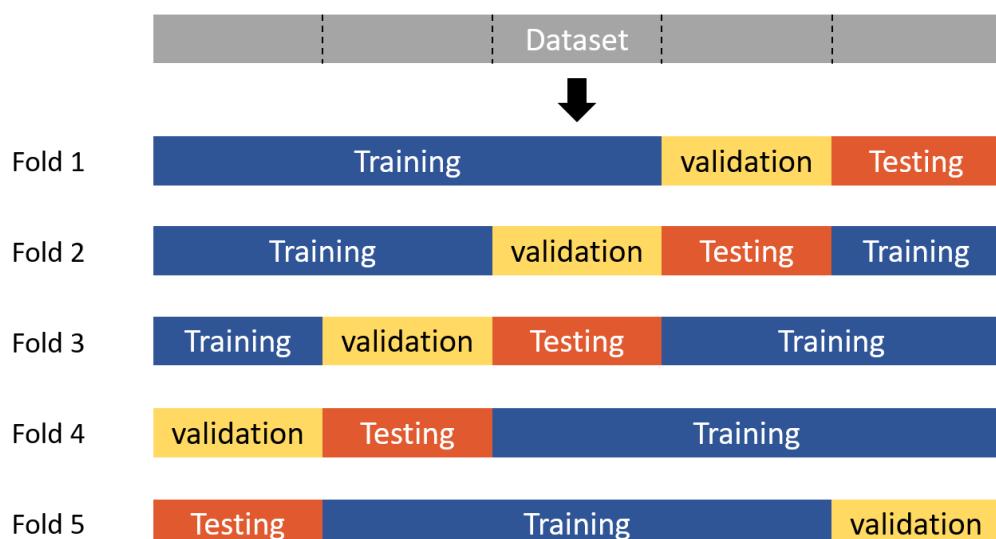


圖 3.19 5-fold cross validation

本研究之原始資料共 250 筆，其中 175 筆為 NonIA (AIS / MIA)、75 筆為 IA，分成五個集合後每一個集合會有 15 筆 NonIA 與 35 筆 IA，在驗證集與測試集樣本各自有 50 筆資料，訓練集資料除了原始的 150 筆影像外，還會從這些原始影像所產生出的增量資料當中抽取 NonIA：IA 比例為 3:7 的部分資料一起進行訓練。下表 3.2 為五個 fold 在第一輪 training 其各自的訓練、驗證、測試的資料數量：

表 3.2 每個 fold 各自的訓練、驗證、測試及資料數量

	Fold1	Fold2	Fold3	Fold4	Fold5
Training	4047	4047	4047	4047	4047
Validation	50	50	50	50	50
Testing	50	50	50	50	50

由於本研究之方法會在第二輪到第五輪的訓練過程中透過資料更新的權重從這 4047 筆資料中以取後放回的方式抽取 1500 筆資料加入訓練集，所以除了第一輪之外，其餘的 training 數量都為 5547。

3.3.2 比較文獻介紹之方法

由於第二章所提到之五篇文獻都與本研究的目的相近且都有相當不錯的數據結果，所以下列將針對第二章提到之文獻利用本研究之資料重複其做法亦或是利用其挑選出來的特徵進行比較，並觀察文獻之樣本資料與本研究資料之差距，探討結果差距的原因。

第一篇文章 2019 年，Fan Li [8] 等人經過一連串特徵挑選過程後選定之特徵為：原始影像計算出的 GLCM 特徵—cluster tendency 與 x,y,z 軸皆做過 low-pass filter 的 wavelet 影像計算之 GLCM 特徵—correlation，由於文中提及這兩個特徵具有一定的分辨能力，對於其不論是在有做過 contrast-enhanced 的 external



validation cohort 或是 non-contrast 的 external/intra-class validation cohort 皆有良好的表現，是相當具有代表性與穩定性的特徵，所以本研究利用這兩種特徵來建立 logistic regression 模型，使用 5-fold cross validation，計算 5 fold 之測試集結果。下圖 3.20 為利用本研究資料之 ROC curve，選擇的 cut-off value 為 0.2799，其計算之正確率、靈敏度、特異度與 AUC 分別為 55.20%、57.33%、54.29%、0.5545，相較於此篇文獻之結果（圖 3.21）正確率為：90.8%、84.0%、88.1%，靈敏度：97.5%、82.2%、85.7%，特異度：94.4%、86.7%、89.8%，AUC：0.971、0.942、0.936 相距甚遠。

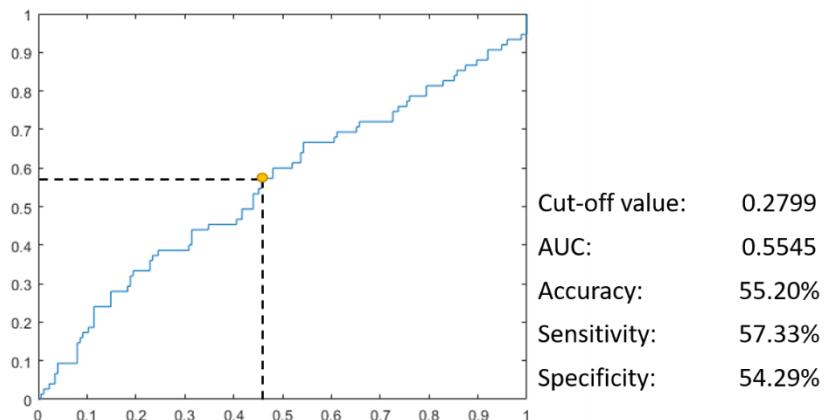


圖 3.20 本研究資料重複 Fan Li [8]等人提倡方法之 ROC curve

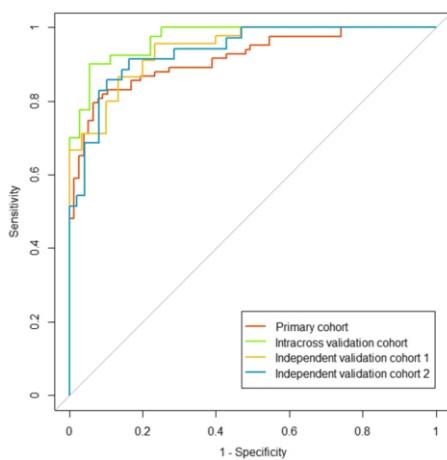


圖 3.21 Fan Li [8]等人提倡方法之 ROC curve

於文獻內，作者提及 Lim 等人[25]計算 IA 之 mean CT 落在-507HU 與其研究相近，於文獻附檔中，也有詳細的兩類 mean CT value，下表 3.3 為作者於



primary cohort、intra-cross validation cohort 及兩個 external validation cohort 兩類別資料的mean CT value ± CT value standard deviation，在 NonIA 的部份其平均值落在-622 HU ~ -586 HU 之間，IA 則落在-398 HU ~ -466HU，在所有 cohort 中兩類平均 CT value 皆有顯著差異，而本研究資料之 IA mean CT value 為-704.3、NonIA mean CT value 為-726.8，如表 3.4，兩者所納入的資料在 mean CT 上並沒有顯著差異，並且其值與文獻提供的數據是有落差的，故可能為造成無法達到文獻這麼好的結果的原因。

表 3.3 Fan Li [8]等人研究資料 mean CT value 比較

	Non-Invasive	Invasive
Prime cohort	-586.4 ± 104.6	-424.7 ± 151.6
Intra-cross validation cohort	-587.7 ± 110.5	-408.4 ± 161.1
External validation cohort 1	-601.9 ± 90.0	-466.4 ± 111.6
External validation cohort 2	-622.1 ± 126.5	-398. ±161.3

表 3.4 本研究資料之 mean CT value

	Non-Invasive	Invasive
My data	-726.8 ± 50.1	-706.3. ±48.3

2020 年，Hwan-ho Cho [9]等人透過 lasso 的方式從 40 個 radiomics features 選取具代表性的五個特徵—Range, GLCM–entropy, ISZM–size zone variability, density, mass，並利用三種分類器：SVM、logistic regression、Random forest 計算其結果。而除了上述之特徵外，作者有提出新的特徵提取方式，也就是文獻探討中所提到的 marginal features，以 2D 切面中腫瘤內部灰階值之 5/100, 35/100, 65/100, 95/100 作為計算的閾值，這組灰階閾值從低到高攀升，計算在每一個灰階閾值下整顆腫瘤灰階低於此閾值之 pixel 數量占整顆腫瘤 pixel 數量的比例繪製成累積分布函數 CDF，繪製完累積分布函數圖後，對這個曲線取斜率再計算出下列四個特徵：平均、標準差、峰度與偏態，即為 marginal features，將前面 lasso 選出之五

個特徵與 marginal features 同時放入三個分類器比較結果。下表 3.5 為作者之結果，其中 Model 1 是利用 lasso 選出的五個特徵建立模型分類結果，而 Model 2 則為有加入 marginal features 所得到的分類結果，可以觀察到三種分類器在 Model 2 的分類效能相對於 Model 1 都有所提升，其中以 logistic regression 的結果最佳，正確率、靈敏度、特異度與 AUC 分別從：89.36%、89.47%、88.89%、0.9766 提升至 91.49%、89.47%、100%、0.9825。文獻中有提及從 lasso 選出之五個特徵中有些與侵犯性是有相關的[26]，所以接下來將針對本研究所用之樣本分成 5-fold cross validation，比較利用五個特徵與加入 marginal features 後分別用不同分類器分類之結果與文獻做比較。由圖 3.22 可以看出在尚未加入 marginal 模型前是以 random forest 的分類結果最好，其正確率、靈敏度、特異度與 AUC 分別為 59.20%、60%、58.86%、0.6221，其結果與文獻差距很大；另一方面不論是利用何種分類器進行分類，再加入了 marginal features 後，全部的分類結果皆退步（圖 3.23），三者的 AUC 都不到 0.6000，所以用本研究之資料在尚未加入本篇研究產生之新特徵前表現結果是較好的。

表 3.5 文獻結果統計表[9]

		Accuracy	Sensitivity	Specificity	AUC	Adjusted R-squared	P value
	Training classifier						
LR	Model 1	0.6613	0.5909	0.7245	0.7490	0.1757	1.5725E-09
	Model 2	0.6828	0.6250	0.7347	0.7507	0.1781	1.1879E-09
SVM	Model 1	0.6935	0.7159	0.6735	0.7767	0.2201	4.8897E-11
	Model 2	0.7419	0.7841	0.7041	0.8291	0.3085	8.1370E-16
RF	Model 1	1.0000	1.0000	1.0000	1.0000	0.9326	6.3055E-110
	Model 2	1.0000	1.0000	1.0000	1.0000	0.9366	2.2911E-112
	Validation						

classifier						
	Model 1	0.8936	0.8947	0.8889	0.9766	0.2806
LR	Model 2	0.9149	0.8947	1.0000	0.9825	0.2700
						1.0832E-04
SVM	Model 1	0.8085	0.7895	0.8889	0.8450	0.1907
	Model 2	0.8511	0.8421	0.8889	0.9240	0.3323
RF	Model 1	0.8298	0.8158	0.8889	0.9211	0.4264
	Model 2	0.8723	0.8684	0.8889	0.9678	0.6078
						6.4906E-11

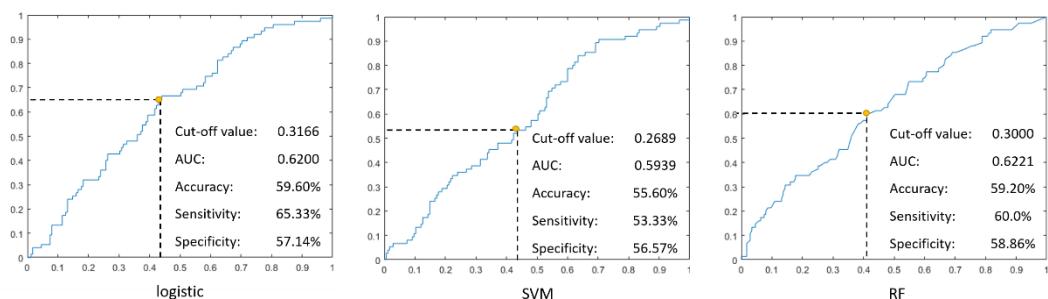


圖 3.22 model 1 產生之結果

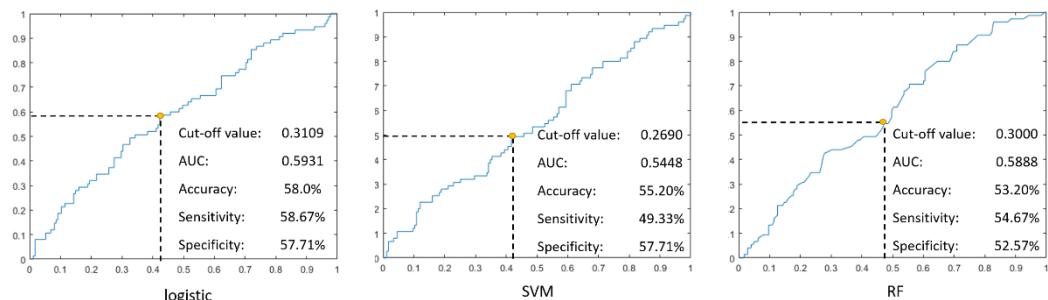


圖 3.23 model 2 產生之結果

從文獻呈現的結果可以發現在 SVM 與 logistic regression 的訓練集結果普遍比 validation 結果差，作者在討論有提及 validation cohort 的資料較為極端，以 IA 為例，validation cohort 的 IA 相較於 training cohort 大，所以在判別上也可以是一個特徵，另外 IA 數量在 validation cohort 也較多，其統計的直徑大小的表格如表 3.6，兩項特徵在訓練及驗證集資料中皆有顯著差異，而本研究全部資料在 IA 的部份直徑 $\leq 1\text{cm}$ 有 43 筆， $1\text{cm} < \text{直徑} \leq 2\text{cm}$ 有 124 筆， $2\text{cm} < \text{直徑} \leq 3\text{cm}$ 有 8 筆，nonIA 中直徑 $\leq 1\text{cm}$ 有 19 筆， $1\text{cm} < \text{直徑} \leq 2\text{cm}$ 有 45 筆， $2\text{cm} < \text{直徑} \leq 3\text{cm}$ 有 11 筆，



並沒有納入超過 3 公分之腫瘤，在大小上如表 3.7 所示沒有太大的差異。

表 3.6 文獻對於訓練及驗證集資料比較表[9]

		Training(n=186)	Validation(n=47)
Tumor size (cm)	≤ 1	31	5
	1 <≤ 2	108	22
	2 <≤ 3	41	13
	> 3	6	7
Histopathology	NonIA	98	9
	IA	88	38

表 3.7 本研究資料於兩類別之最大直徑分布

		IA(n=75)	Non-IA(n=175)
Tumor size (cm)	≤ 1	43	19
	1 <≤ 2	124	45
	2 <≤ 3	8	11
	> 3	0	0

同年 Linyu Wu [10]等人，利用腫瘤本體與腫瘤本體膨脹 2mm 之範圍做特徵提取，提取的特徵分別有 histogram, form-factor, GLCM, GLRLM。產生完特徵後資料會透過 SMOTE 的方式將兩類資料調成接近 1:1 的比例，再進行特徵挑選：mRMR 與 LASSO，當 LASSO 做完會產生每個特徵對應的係數，選取出係數不為 0 之特徵，將特徵值與係數相乘後全部相加即可得到「radiomics score」最後透過此分數通過 logistic regression 得到最後的分類結果。文獻比較三種不同的 radiomics score，其中一個為僅用腫瘤本體產生之特徵執行上述的流程得到的 intra-nodular radiomics score 以及僅用腫瘤本體往外膨脹 2mm 範圍所產生之特徵執行上述流程得到之 peri-nodular radiomics score，與兩種特徵同時執行得到之 gross radiomics score，其分別的正確率為：82.9%、77.1%、88.6%，靈敏度：90%、80%、80%，



特異度：80%、76%、92%，及 AUC 為：0.852、0.820、0.876，其中以第三種 gross radiomics score 得到的結果為最佳。圖 3.24 為 training cohort 與 validation cohort 兩者計算出的 gross radiomics score 的 boxplot，由圖可觀察到不論在哪一個 cohort 下 label 為 1 (IA) 的值大抵來說都比 label 為 0 (AIS/MIA) 的大。後面一樣將針對本研究之資料提取 GLCM、GLRLM、GLSZM 以及 Hostogram 特徵以重複文獻作法同樣通過 5-fold cross validation 的方式取得其分類結果，並繪製出五個 fold 在 training 與 validation 的箱型分布圖比對兩個類別的分布狀況，見圖 3.25、圖 3.26、圖 3.27。

僅用腫瘤本體產生之特徵重複文獻做法在 5-fold 平均的正確率、靈敏度、特異度與 AUC 分別為：57.60%、64%、54.86%、0.5928，而僅用腫瘤外部產生之特徵重複文獻做法在 5-fold 平均的正確率、靈敏度、特異度與 AUC 分別為：52.80%、58.67%、50.29%、0.5493，最後同時使用腫瘤內外部特徵所得到的結果為：58.0%、58.67%、57.71%、0.5972，從三者結果來看同時使兩者之 AUC 為最佳：0.5972，而單獨使用腫瘤外特徵得到之結果最差：0.5493。

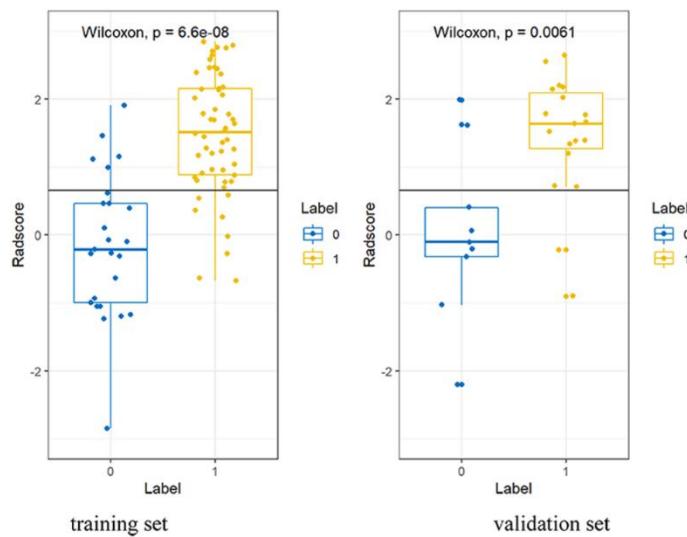
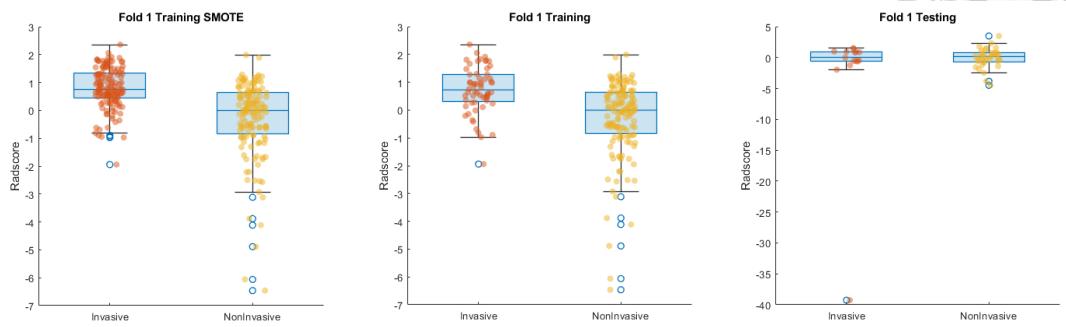
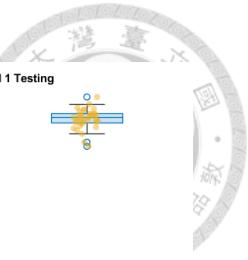
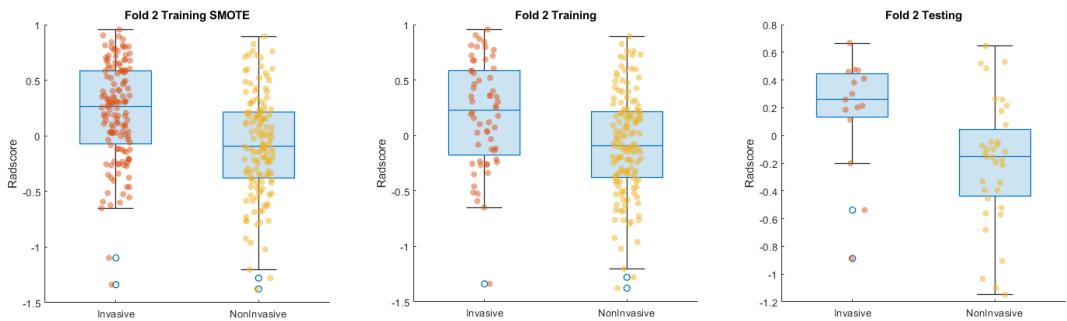


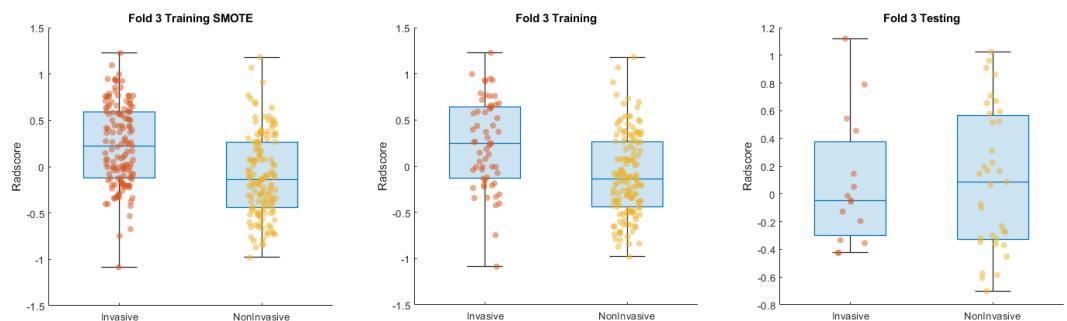
圖 3.24 文獻 gross radimoics score 於訓練集與測試集之箱型圖[10]



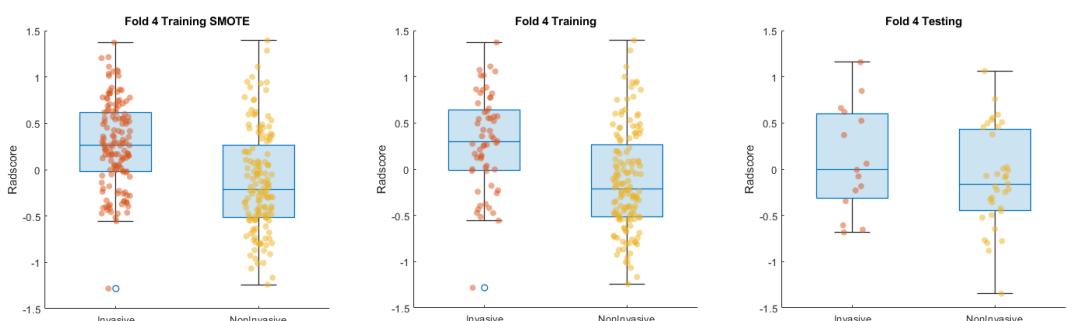
(a) Fold 1



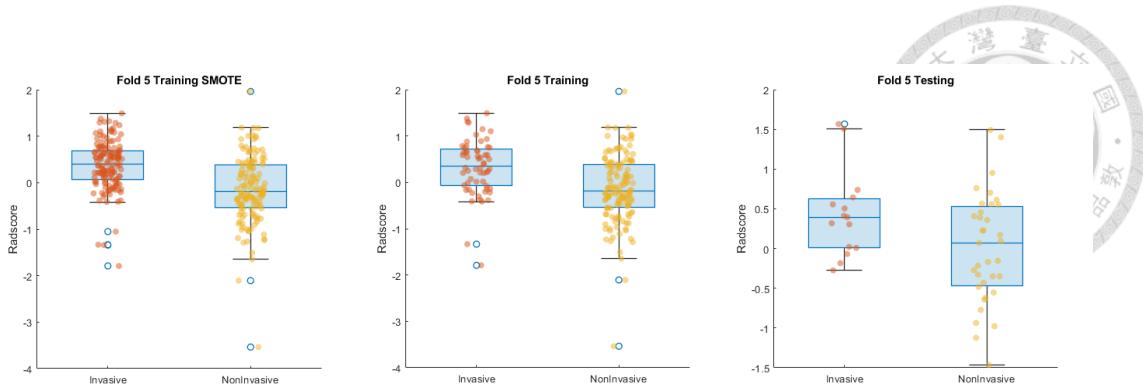
(b) Fold 2



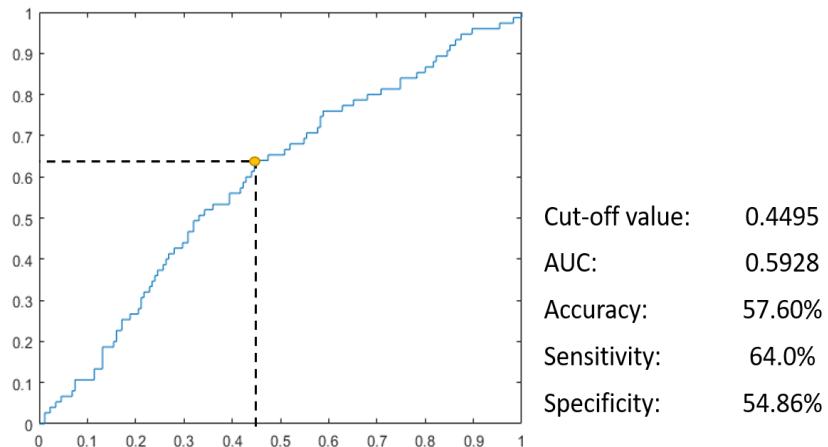
(c) Fold 3



(d) Fold 4

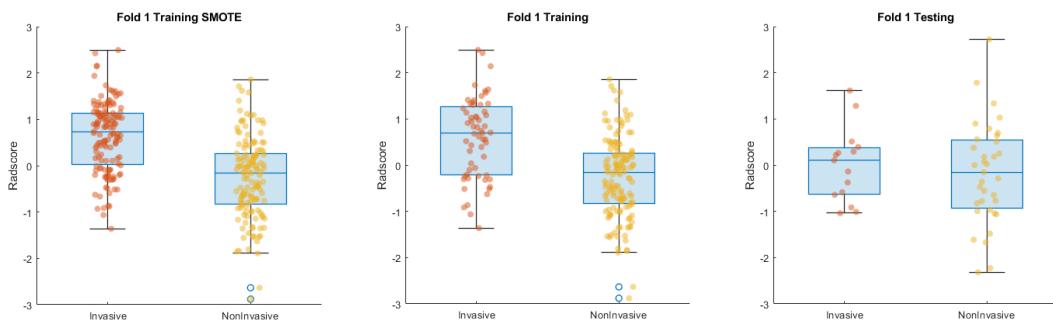


(e) Fold 5

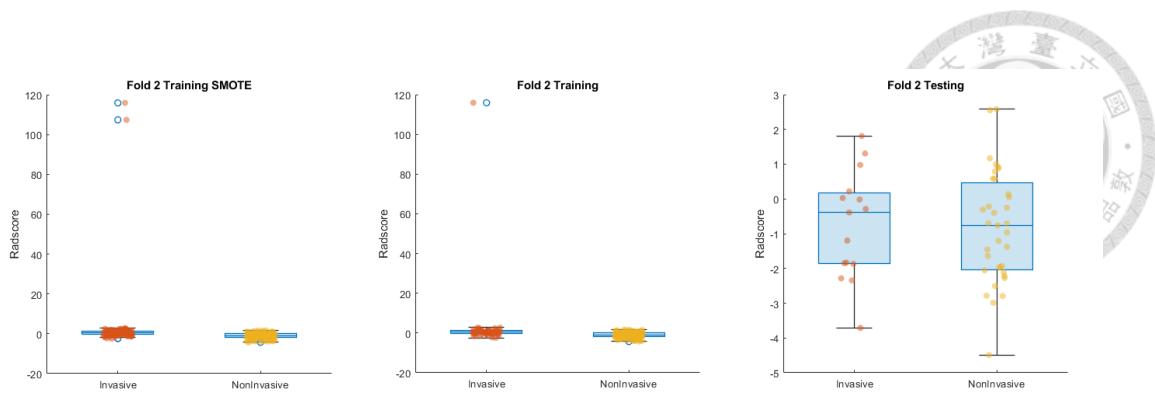


(f) ROC

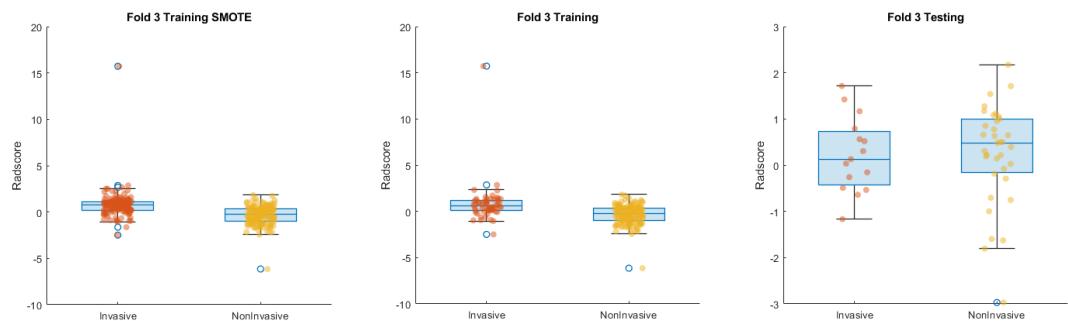
圖 3.25 (a)~(e)為利用腫瘤內部之特徵重複文獻做法其 Fold 1 至 Fold 5 分別在做完 SMOTE 之 training data 兩類分布箱型圖、原始 training data 兩類分布箱型圖、testing data 兩類分布箱型圖(f)為 5-fold 之 ROC



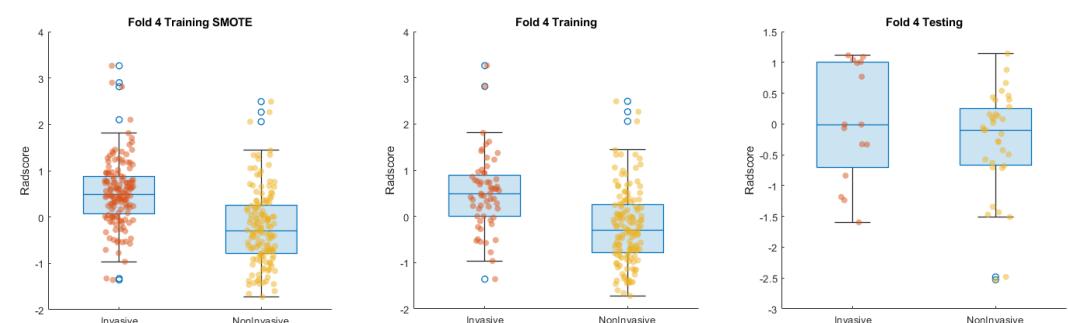
(a) Fold 1



(b) Fold 2



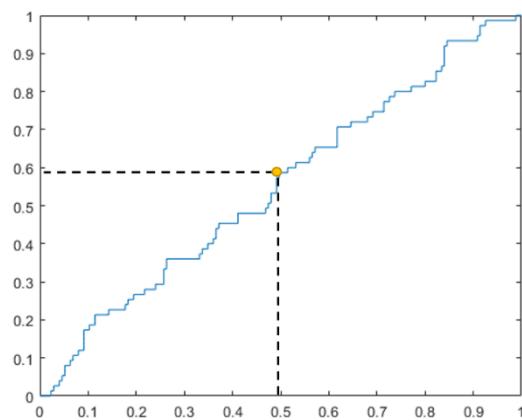
(c) Fold 3



(d) Fold 4



(e) Fold 5



(f) ROC

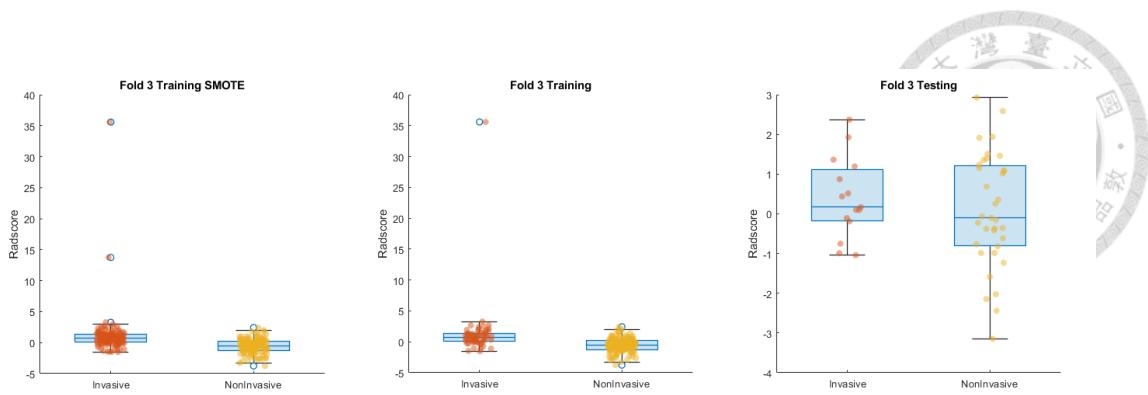
圖 3.26 (a)~(e)為利用腫瘤外部之特徵重複文獻做法其 Fold 1 至 Fold 5 分別在做完 SMOTE 之 training data 兩類分布箱型圖、原始 training data 兩類分布箱型圖、testing data 兩類分布箱型圖(f)為 5-fold 之 ROC



(a) Fold 1



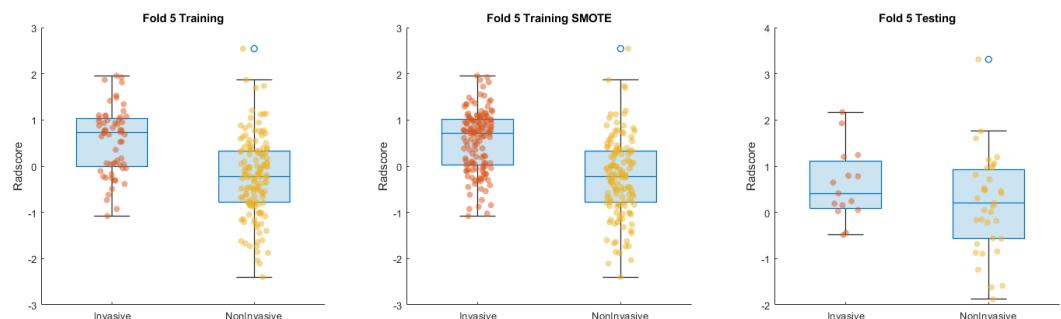
(b) Fold 2



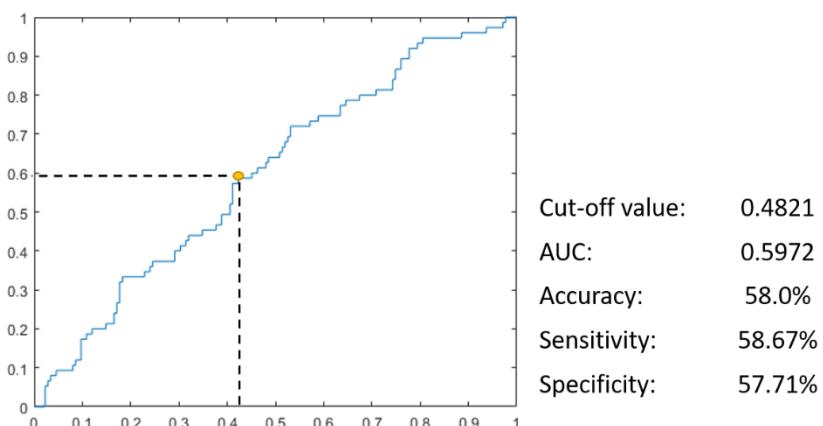
(c) Fold 3



(d) Fold 4



(e) Fold 5



(f) ROC

圖 3.27 (a)~(e)為利用腫瘤外部與內部之特徵重複文獻做法其 Fold 1 至 Fold 5 分別在做完 SMOTE 之 training data 兩類分布箱型圖、原始 training data 兩類分布箱型圖、testing data 兩類分布箱型圖
(f)為 5-fold 之 ROC

於文中，作者提到直徑對於其分類效果也是有幫助的，在 gross radiomic signature 加入最大直徑進行分類在 training 的部分是有提升的 (AUC 從 0.869 到 0.917)，在 validation 的結果則與沒加入最大直徑的結果差不多，AUC 都為 0.876。本研究使用之資料在 IA 平均直徑為 14.2 ± 5.1 (mm)，NonIA 則為 12.6 ± 3.8 (mm)，而文中的 training set IA 平均直徑： 14.81 ± 5.86 、NonIA 為： 11.72 ± 6.13 以及 validation set IA: 14.62 ± 4.88 、NonIA : 9.36 ± 4.11 ，相較本研究資料差距為大，並且在本研究資料之 nonIA 其大小與文獻資料相比較大，而通常腫瘤越大越有機會為 IA，是一項對分類有幫助的因素[27-29]，所以可能由於其大小相近，少了一個分類依據導致結果沒這麼好。

2020 年，Guangyao Wu[11]等人，在分類是否具侵犯性的腫瘤分類上採取的方式是設定灰階閾值-50HU，將所有腫瘤區域分成兩個部份分別為 Ground glass region (-50HU 以下區域) 以及 solid region (-50HU 以上區域)，透過 RadiomiX Discovery Toolbox，提取兩區域的特徵，提取完後進行特徵選取，並經過一連串的挑選後，最後從 solid part 的部分挑選出三個特徵：(1) 影像於 y,z 軸做 high-pass filter 及 x 軸做 low-pass filter 後，從 NGTDM 提取的 contrast 特徵，(2) 影像於 x 軸做 high-pass filter 及 y,z 軸做 low-pass filter 後，從 GLDZM 提取的 DZV 特徵，以及 (3) 影像於 y 軸做 high-pass filter 及 x,z 軸做 low-pass filter 後，從 GLDZM 提取的 DZV 特徵；而 ground glass region 則挑選出四個特徵：(1) GLRLM 的 SRLGE，(2) 影像於 x,y,z 軸做 low-pass filter 後，得到位於 10^{th} percentile 的灰階值，(3) compacteness2 以及 (4) 影像於 x,y,z 軸做 low-pass filter 後，計算的 local intensity peak，以上總共七個特徵，得到之分類結果正確率、靈敏度、特異度與 AUC 分別為：93%、98%、78%、0.9800。作者將此結果分別與 clinical model、volumetric model 以及使用不同範圍提取之 radiomics feature 建立



的模型進行效果比對，其效果比對如表 3.8，

表 3.8 文獻不同模型結果比較[11]

Model	AUC	Accuracy	Sensitivity	Specificity
Clinical-sematic	0.90	0.74	0.66	0.94
volumetric				
GTV	0.87	0.74	0.68	0.89
GG&S	0.90	0.74	0.64	1
Radiomics				
GTV	0.88	0.72	0.64	0.94
Ground glass region	0.96	0.9	0.88	0.94
Solid region	0.82	0.75	0.72	0.83
Perinodular region	0.66	0.54	0.54	0.56
GG&S	0.98	0.93	0.98	0.78
GTV&P	0.90	0.72	0.64	0.94
GG&S&P	0.95	0.88	0.90	0.93

其中 GTV 為原始腫瘤邊界；GG 為腫瘤-50HU 以下的區域、S 為腫瘤-50HU 以上的區域；P 為腫瘤往外膨脹 5mm 的區域。

接下來將利用此篇文獻提出之 GG&S volumetric model 以及 ground glass lesion 以及 solid region 挑出來的七個特徵建立的 radiomics model，對本研究資料進行 5-fold cross validation 的模型建構，與文獻之結果做比較，而由於作者所使用之閾值太高導致本研究所使用的資料無法分離出 solid region 的部分，故改以使用在文獻上較常使用的閾值：-190HU、-290HU 做代替；另外除了使用本研究限制 solid ratio<0.25 之資料進行訓練外，也會加入先前經醫師挑選後認為 solid ratio>0.25 之資料做訓練，觀察其結果，如下表 3.9：



表 3.9 本研究資料重現文獻之方法結果

Model	AUC	Accuracy	Sensitivity	Specificity
volumetric				
GG&S(-190HU)	0.5151	0.4804	0.5200	0.4686
GG&S(-290HU)	0.5157	0.5080	0.5333	0.4971
GG&S(-190HU) *	0.5659	0.5335	0.5147	0.5459
GG&S(-290HU) *	0.5742	0.5820	0.6176	0.5556
Radiomics				
GG&S(-190HU)	0.5938	0.6000	0.6000	0.6000
GG&S(-290HU)	0.6039	0.6160	0.5733	0.6343
GG&S(-190HU)*	0.5577	0.6114	0.6047	0.6133
GG&S(-290HU)*	0.5843	0.5648	0.6579	0.5467

表中*代表加入 solid ratio>0.25 之資料結果

從表 3.9 結果來看，不論是以-190HU 還是-290HU 當作分界，在 solid ratio<0.25 之資料以及未受限制資料所做出來的結果都與文獻差距懸殊，故後面會對文獻資料以及本研究資料進行探討，分別對於 solid portion 以及平均 HU 值方面進行著手。

在文中，作者有提供樣本資料的 solid portion 以及 2D attenuation，整理文獻資料及本研究資料統整表如下表 3.10：

表 3.10 本研究資料與文獻資料於 solid portion 及 attenuation 統整表

		Solid portion	attenuation
Paper's data	IA	43%	-293.2
	Non-IA	61%	-448.9
Our research's data	IA	<25%	-704.3
	Non-IA	<25%	-726.8

從表 3.10 中可以看出文獻資料利用閾值-50HU 分割出的 solid part 佔比在兩類中



皆>40%，可見其研究資料在 IA 與 non-IA 比較上相較於本研究資料實質部分佔的比重較重，另外從 attenuation 這個特徵值來看，可以觀察到 IA 相較下是亮許多的，此特徵值也能當作一個判斷是否具侵犯性的準則。由於本篇文獻著重於分析 part solid 之腫瘤，利用 ground glass region 與 solid region 分別提取特徵去建立模型，其資料皆為 part solid GGN，而不存在 pure GGN 的資料，故在資料分布上，有些不同，也造成兩者結果的差異，再者，相較於本研究為希望從 solid ratio 較低之資料進行分類，故在限制上有差異。

2020 年 Jing Gong [12]等人，透過深度學習的方式進行 IA 與 non-IA 的分類，其建構之模型架構主要是基於 residual learning-based CNN network 來進行修改，比較特別的地方是模型放入的 input 不是拿涵蓋腫瘤本體的 3D VOI 進行訓練，而是將每筆資料從 coronal view, sagittal view 以及 axial view 各切出一張涵蓋腫瘤的 64*64 region of interest，並將這三張 2D 的 ROI 堆疊形成一張 RGB 影像，做完資料增量後即進行模型訓練。文獻提出的方法僅訓練一個模型並用 102 筆 external validation cohort 當作測試集來檢視模型訓練的效果，其 AUC 可以達到 0.92 ± 0.03 的表現，正確率、F1-score、weighted average F1-score 分別為：83.3%、74.6%、83.3%。此外，文中提及實驗的測試數據也同時分別由具有 15 年經驗與 6 年經驗的放射科醫師依據 solid portion 來進行人為的判斷，總共分為五類，最後判定結果會是 rating 為 1~3 之資料歸類為 non-IA，而 4、5 為 IA，從醫師判斷數據表現出較資深的放射科醫師做出的判斷與 AI 做出的判斷力是相當的，為 AUC： 0.92 ± 0.03 ，而另一位放射科醫師的結果為 AUC： 0.90 ± 0.03 ，三者 ROC 如圖 3.28。為重現文獻作法，會將本研究之資料分成一組 training、validation、testing data，利用文獻提及之模型進行一次模型訓練，訓練結果所得之正確率、靈敏度、特異度與 AUC 為： 0.5829 ± 0.04 、 0.52 ± 0.03 、 0.5640 ± 0.03 與 0.5749 ± 0.01 ，結果與文獻的預測效能有所落差，如圖 3.29。

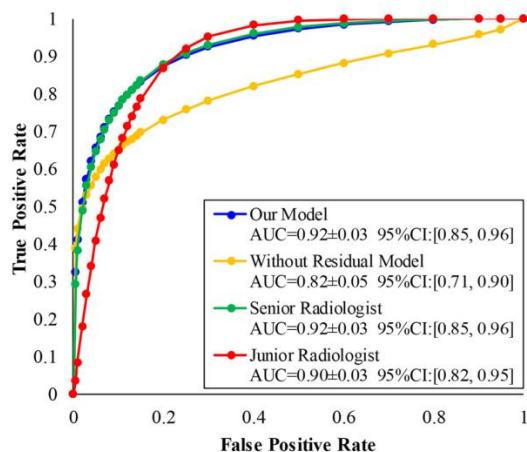


圖 3.28 模型與醫師判斷之 ROC 曲線[11]

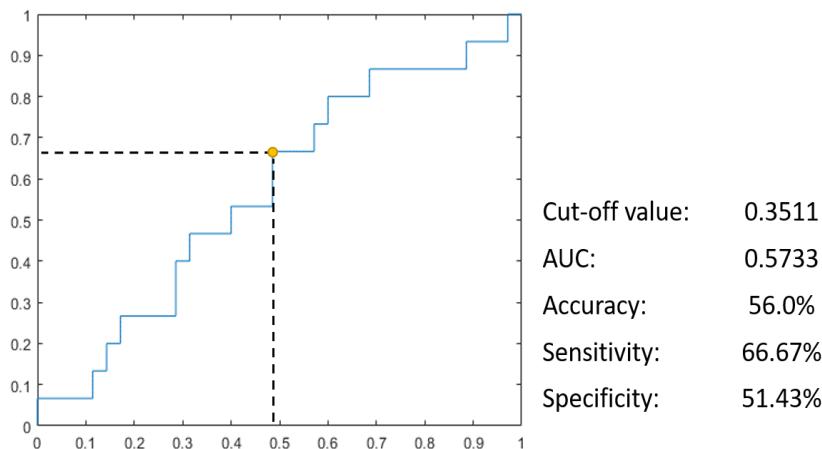


圖 3.29 本研究測試資料之 ROC 曲線

由文中放射科醫師給予腫瘤 solid portion 所占百分比之表 3.11 來看，其資料只要不是整顆都為 solid part 就可以納入，而放射科醫師單純透過判斷 solid portion 所占百分比即可達到 $AUC = 0.92 \pm 0.03$ 與 0.90 ± 0.03 的好結果，表示 solid part 對於判斷腫瘤為 IA 或是 NonIA 是一項非常重要的指標，但於本研究對 solid ratio 有所限制，限制於 $\text{solid ratio} < 0.25$ ，相較於文獻之資料較難作區別。

表 3.11 放射科醫師對 GGN 分類標準[12]

Percentage of solid portion	Likelihood of GGN being IA	Rating
$\leq 20\%$	Highly unlikely	1
(20%, 40%]	Moderately unlikely	2

(40%, 60%]	Indeterminate	3
(60%, 80%]	Moderately suspicious	4
> 80%	Highly suspicious	5

最後一篇 Xianwu Xia[13]等人，融合了放射體學與深度學習方法來進行是否具有侵犯性的分類，首先先建立 RRCNN 對腫瘤進行分割，而訓練好的分割模型會保留前面 feature extraction 的架構與權重並在最後加入兩層全連階層變成分類模型，放入欲分類之資料進行訓練，最後放入測試資料得其預測結果；另一方面，由前面 RRCNN 建立之分割模型分割研究資料得到腫瘤範圍，根據其分割範圍對腫瘤內部進行特徵提取，得到 laplacian of Gaussian、小波轉換、一階（直方圖）、形狀、紋理（包含 GLCM、GLDM、GLRLM、GLSZM）特徵，再利用 ANOVA F-value 的方式做單變量特徵挑選，挑選前 20 個放入 SVM 進行分類。建立好 DL model 與 radiomics model 後就可以對兩者的預測結果做融合，而其最佳融合方式為選擇兩者最高之預測結果當作最後的預測機率值：final prediction = max(radiomics model prediction, DL model prediction)，AUC : 0.90 ± 0.03 ，正確率：80.3%，而單純使用 radiomics 跟單純使用 deep learning 的 AUC 結果分別為： 0.87 ± 0.04 、 0.83 ± 0.05 ，另外，其結果也與兩位放射科醫師做比較，其中一位的正確率為 67.7%，另一位為 70.9%。圖 3.30 為 ROC curve：

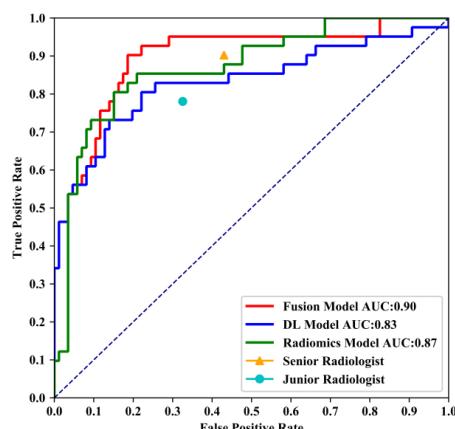


圖 3.30 Radiomics 模型與深度學習模型以及 fusion model 之 ROC 曲線[13]

再來利用本研究之資料重複文獻做法，同文獻一樣使用公開資料庫 LIDC-



IDRI 對分割模型做訓練，再將其訓練好的權重 transfer 到分類的模型中，並使用 5-fold cross validation 進行 fine-tune，在提取腫瘤內部特徵方面，由於本研究所使用的資料本來就有腫瘤的分割圖，所以將直接使用原本的分割結果進行特徵的提取。在 radiomics 的部分，模型分類的預測正確率、靈敏度、特異度與 AUC 分別為：62%、64%、61.14%、0.6143；在深度學習的部分，單獨測試的結果正確率、靈敏度、特異度與 AUC 分別為：58.40%、61.33%、57.14%、0.6089；另外，也仿照了文獻利用不同比例融合深度學習模型與 radiomics 模型預測之結果，得到最高 AUC 之比例為：最終預測值 = 深度模型預測值 * 0.4 + radiomics 模型預測值 * 0.6，其正確率、靈敏度、特異度與 AUC 分別為：58.80%、68.0%、54.86%、0.6318，下圖 3.31 為三者 ROC curve：

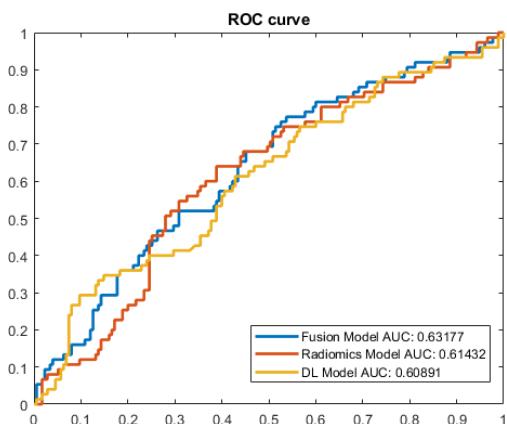


圖 3.31 本研究資料 Radiomics 模型與深度學習模型以及 fusion model 之 ROC 曲線
在文內，作者有提到其納入資料之 mean CT value，在 training dataset 中 IA 與 non-IA 的 mean CT value 為 -439 ± 138 與 -533 ± 116 ，在 testing dataset 則分別為 -381 ± 182 與 -553 ± 142 ，如圖 3.32。相較於本研究資料（圖 3.33）來說屬於 solid part 較多的樣本，故可能為本研究利用本篇文獻之方法無法得到相近之結果的原因。

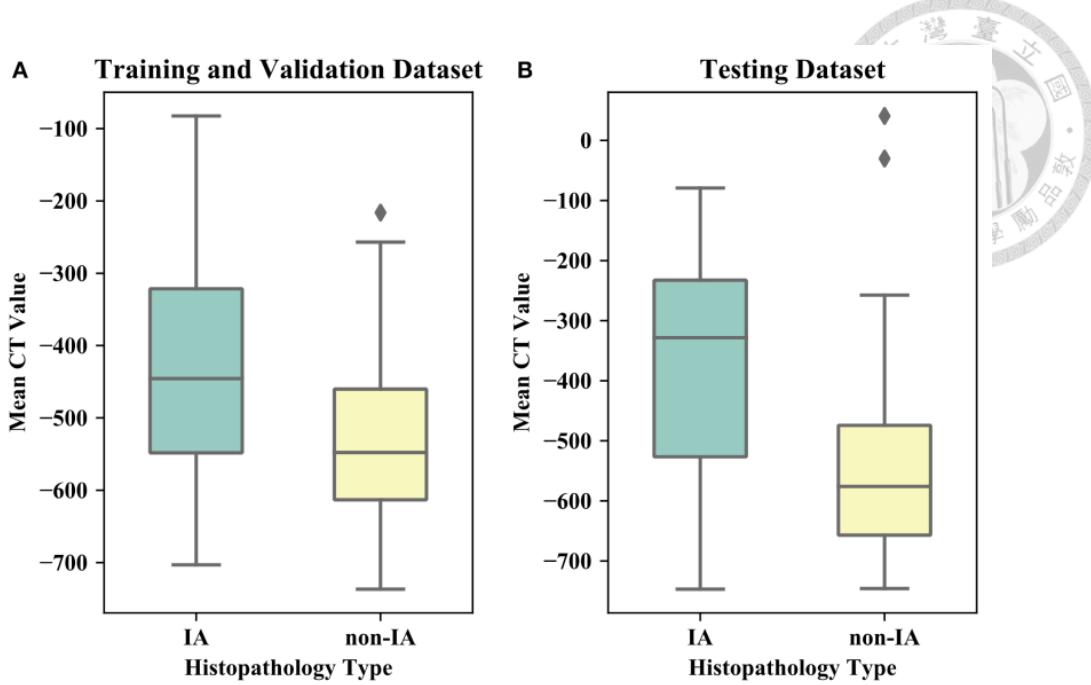


圖 3.32 (a) testing 以及 validation mean CT value (b) testing data mean CT value 箱型圖[13]

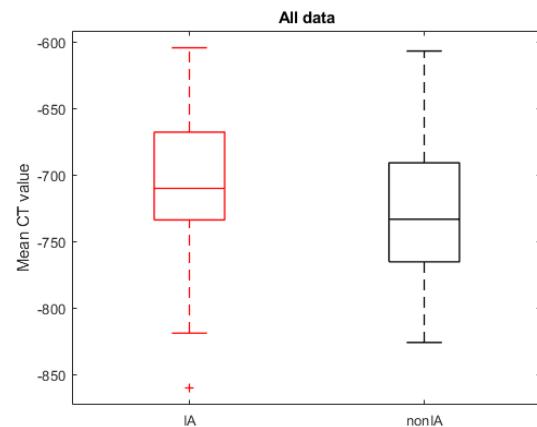


圖 3.33 本研究資料 mean CT value 箱型圖

另外，由於此篇文獻是直接利用 testing data 去找最佳 AUC 以挑選最好的 fusion 方式，但鑑於此種方式有球員兼裁判的問題，故本研究將 radiomics model 也拆分成 training, validation, testing data，並透過 radiomics model 與 deep learning model 中的 validation data 選擇融合方式，再套用至 testing data。radiomics 模型之訓練資料變更後的測試集結果其正確率、靈敏度、特異度與 AUC 分別為：55.20%、58.67%、53.71%、0.5566，於 validation data 中獲得與深度學習模型最佳融合方式為：最終預測值 = 深度模型預測值 * 0.2 + radiomics 模型預測值 * 0.8，其正確率、靈敏度、特異度與 AUC 分別為：60.0%、65.33%、57.71%、0.6376，



將此融合方式套用至 testing data 得到之結果為 56.0%、61.33%、53.71%、0.5927，與直接利用 testing 方法作融合的結果就有落差，所以若直接拿 testing data 所得到的結果的確會提升它的 performance，ROC 曲線如圖 3.34，所以為排除球員兼裁判的疑慮，應使用 validation data 進行融合方式的挑選，再套用至 testing data。

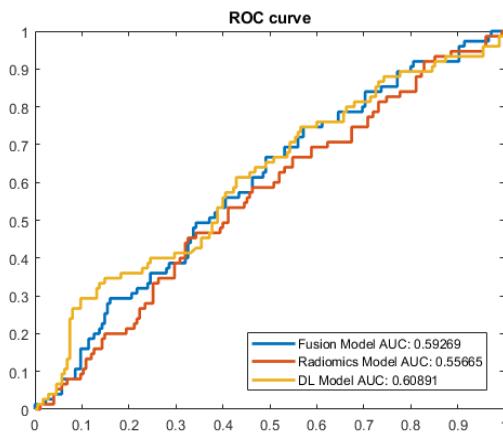
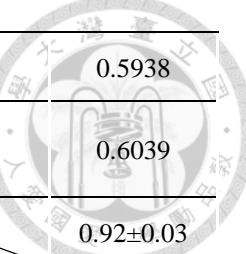


圖 3.34 調整過後的三個模型 ROC 曲線

最後總結所有文獻其結果與利用本研究之資料重現所有文獻做法之結果，統計表如下：

表 3.12 文獻結果與本研究重現文獻做法之結果統計表

		Accuracy	Sensitivity	Specificity	AUC
Fan Li [8]	文獻	90.8%	97.5%	94.4%	0.971
	文獻	84%	82.2%	86.7%	0.942
	本研究	88.1%	85.7%	89.8%	0.936
Hwan-ho Cho [9]	本研究	55.2%	57.33%	54.29%	0.5545
	文獻	91.49%	89.47%	100%	0.9825
Linyu Wu [10]	本研究	59.20%	60%	58.86%	0.6221
	文獻	88.6%	80%	92%	0.876
Guangyao	本研究	58%	58.67%	57.71%	0.5972
	文獻	93%	98%	78%	0.98



Wu[11]	本研究 (-190H 上; -290HU 下)	60%	60%	60%	0.5938
Jing Gong [12]	文献	83.3%			0.92±0.03
Xianwa Xia [13]	文献	80.3%			0.9±0.03
	本研究	56.4%±3.3%			0.5749±0.09
	本研究	58.40%			0.6089

3.3.3 本研究方法之比較

本章節將先前曾經在 radiomics 或者是深度學習方法嘗試過的一些模型進行效能比較，呈現先前嘗試的其他方法在調整參數與模型之下得到之極限。

3.3.3.1. Genetic CNN

此方法是由基因演算法 (Genetic Algorithm) [30]的概念而來，基因演算法來自於生物學中的進化生物學的現象，解釋生物在適應環境過程中會因為其某種優勢或劣勢導致生存能力發生差異，有時生物會透過突變、交配等等的方式來產生新的遺傳因子，在物競天擇下，就會有「適者生存，不是者淘汰」的現象發生，讓適合繼續在這環境下生長的物種持續在這個不斷變化的環境中進化。

先前將基因演算法的宗旨套用在深度學習的模型上，由於即便是相同模型，在訓練過程中所學習到的特徵仍會有差異，所以此方法將相同架構之模型訓練 10 次，得到十個訓練好的模型結果 (模型權重)，再利用 validation data 計算模型的 loss 並將這十個模型進行排序，排序後留下最佳的前四個權重透過 transfer learning 的方式對這四個模型繼續做訓練，而剩下的六個中有兩個會重新初始化參數重新做訓練，另外四個則做 crossover，在 crossover 的部分有做兩種嘗試，一種是結合權重的方式、另一個是平均權重。結合權重的作法是將模型中的四個 convolution kernel



都拆解成前半與後半，也就是當 kernel number 為 4，feature map 大小為 32 時，convolution kernel 就是 $32 \times 32 \times 32 \times 4$ ，而拆解成前半與後半即得到兩個 $32 \times 32 \times 32 \times 2$ ，並保留前半，後半則與其他模型的後半進行結合；平均權重則是將其中一個模型的 convolution kernel 與其他的做平均，得到新的權重，圖 3.35、圖 3.36 為兩種 cross over 的方式。十個模型都更新完參數後就會進到第二次的訓練，在此稱每個訓練為一次的 generation 如圖 3.37，當訓練完五個 generation 後就會從這幾個 generation 中選出 validation loss 最小的當作這個 fold 的模型，並放入測試集數據看結果。

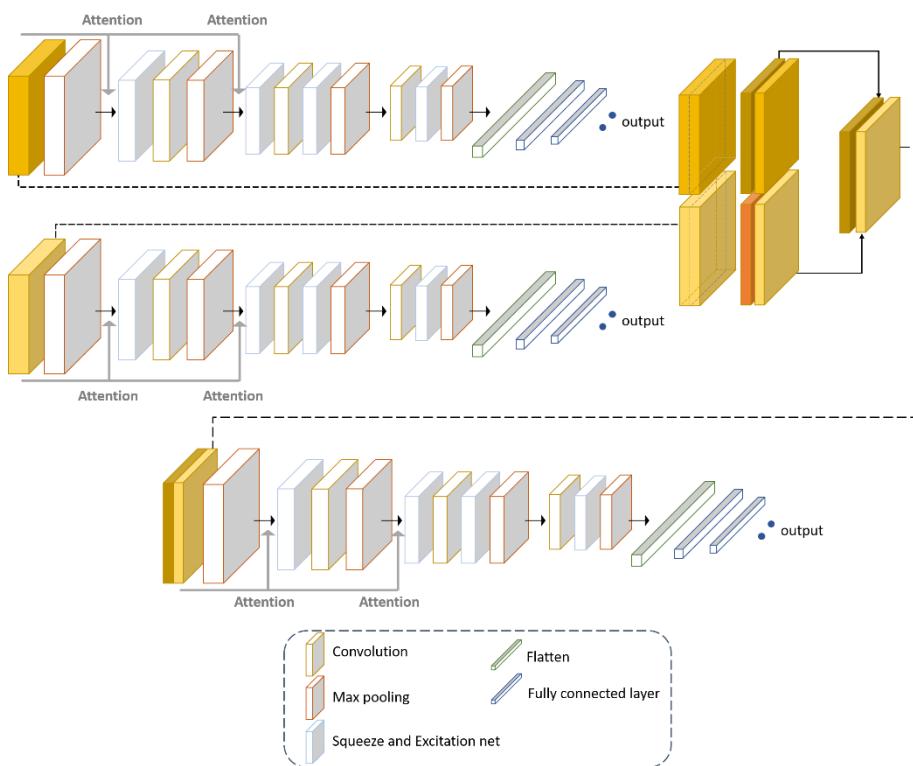


圖 3.35 Combine 的方式，本圖為示意圖，僅呈現其中一個 convolution kernel 之 combine 方式，在其他三個 convolution kernel 會用一樣的方式進行結合

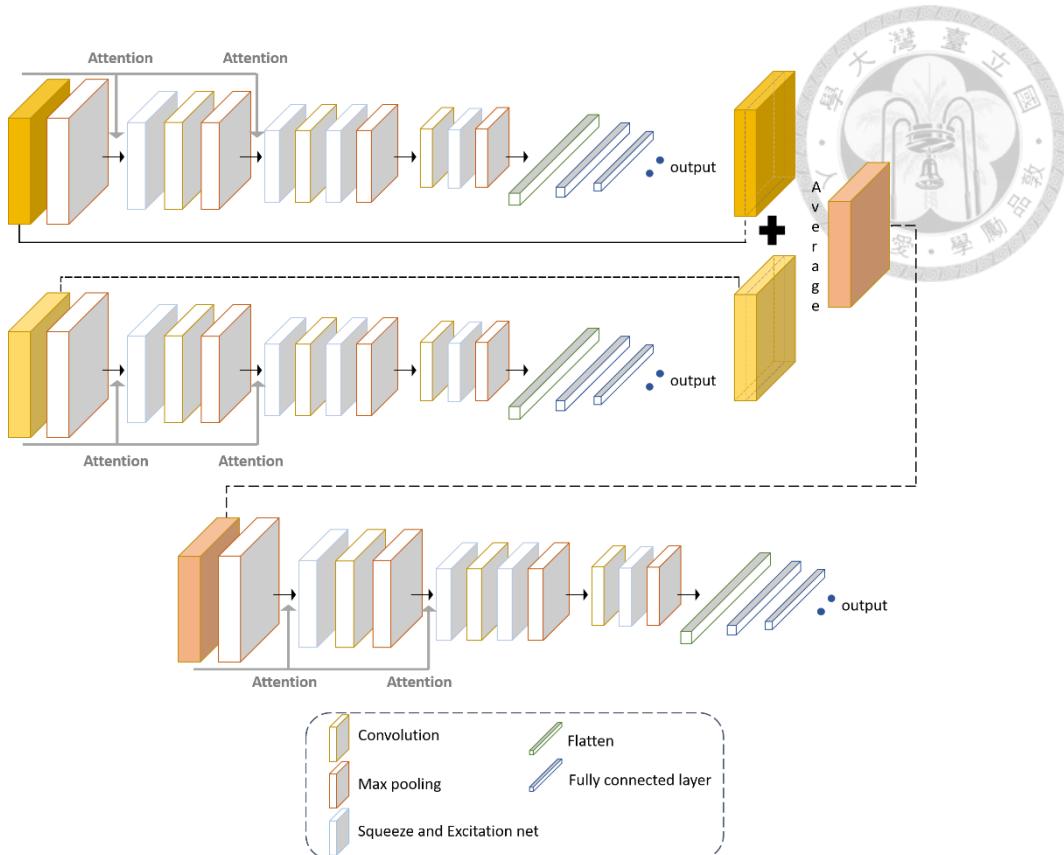


圖 3.36 Average 的方式，本圖為示意圖，僅呈現其中一個 convolution kernel 之 average 方式，在其他三個 convolution kernel 會用一樣的方式進行權重平均

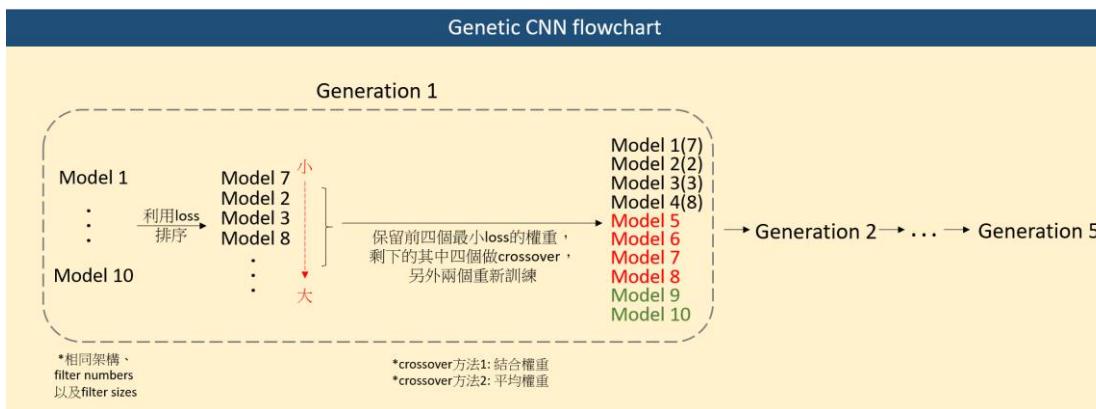


圖 3.37 geneticCNN

在平均權重的部分有嘗試用不同的 filter numbers 看其結果，其中測試了 filter number 為 4、8、16 比較對模型效能的影響，表 3.13 為三者訓練完的結果，從表中可以觀察到在 kernel number 比較小時，模型能在迭代中不斷進步得到相較於前面更小的 validation loss 的結果，最後所選到的 generation 都屬於比較後面的，而 kernel number 為 8, 16 的比較傾向選到較前面的 generation，也就表示後面演化出來的模型參數並沒有比較好的訓練結果，一方面也推斷當模型參數量越大，對於



資料數量較少的 training set 來說是比較難去訓練，並且容易過擬化。在三種 kernel number 的情況下，其 AUC 分別為：0.6329、0.5745、0.6000。

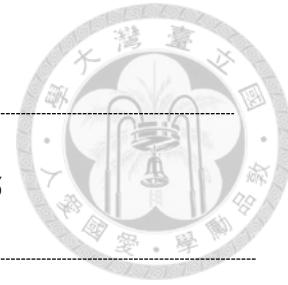
表 3.13 平均權重之效能統計表

Kernel number	4	8	16
Generation/Model			
Fold 1	5 / 5	2 / 2	1 / 4
Fold 2	5 / 2	1 / 5	1 / 7
Generation/Model			
Fold 3	5 / 3	4 / 2	2 / 5
Generation/Model			
Fold 4	5 / 5	4 / 1	2 / 1
Generation/Model			
Fold 5	4 / 5	1 / 3	2 / 8
Generation/Model			
Accuracy	62%	62%	57.20%
Sensitivity	66.67%	58.67%	60%
Specificity	60%	63.43%	56%
AUC	0.6329	0.5745	0.6000

表 3.14 利用結合權重進行 crossover 並將 kernel number 設定為 4 所得到之結果，其效果相較於平均權重來看較為不顯著，在 fold 2 與 fold 5 尤其訓練不彰，只選到 generation 2 與 1，對於後面的訓練並沒有更好的效果，AUC 方面僅有 0.5432。

表 3.14 結合權重之效能統計表

Kernel number	4
Fold 1	4 / 2



Generation/Model

Fold 2

2 / 5

Generation/Model

Fold 3

5 / 8

Generation/Model

Fold 4

5 / 1

Generation/Model

Fold 5

1 / 2

Generation/Model

Accuracy	56.40%
Sensitivity	49.33%
Specificity	59.43%
AUC	0.5432

3.3.3.2. 結合多個 CNN model dense 層之 feature 進入 ML

這個方法的想法是來自於即便是同樣的模型，在每次訓練的過程當中仍可能因為找到不一樣的 local minimum 而產生不同的特徵，所以透過多次 training 模型後收集所有出現在模型最後一層 dense 之結果當作新的特徵，結合這些特徵後挑掉相關性較強($\text{correlation coefficient} \geq 95\%$)的即可生成一組全新特徵，再透過一些特徵挑選、合併的方式選出較具代表性的進入分類器進行分類，如圖 3.38。在挑選模型時，有兩種不同挑選方式，一種是以資料答錯程度較低的方式也就是選取 validation loss 較低的來測試 testing set，另一種是選擇三項評估效能皆高的方式也就是 validation data 算出 sensitivity + Specificity + accuracy 總和最高的那組權重來測試 testing set。在第一種方式結果最好的為將模型產生出的特徵透過 Forward selection 演算法選出最多十個，再通過 Principle component analysis(PCA)



選出前三個特徵當作最後的特徵進入貝氏分類器，其正確率、靈敏度、特異度與 AUC 分別為 57.60%、60.0%、56.57% 與 0.6072；而第二種方式產生之特徵一樣先透過 forward selection 選出十個，再利用窮舉法組合出最佳的三個特徵，放入 SVM 得到之正確率、靈敏度、特異度與 AUC 分別為 59.60%、58.67%、60.0% 與 0.6010。

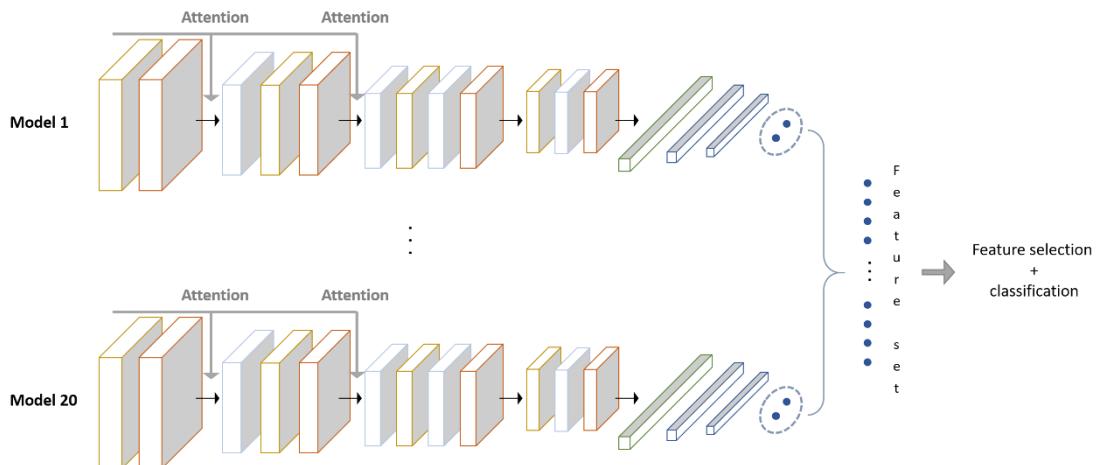


圖 3.38 結合模型 dense 層特徵做特徵挑選與分類

3.3.3.3. Pre-train model

由於本研究之資料僅有 250 筆，對於 data driven 的深度學習來說資料量較為不足，而在資料有限的情況下除了做資料增量製造多一點訓練樣本外，還可以透過 pretrain 的方式先利用別的資料對模型進行訓練，訓練完再使用目標資料進行 finetune。最常被拿來當作 pretrain 資料就是 image net[31]，利用多張自然影像訓練模型，讓其學得紋理與一些深層的資訊，唯一的缺點是 image net 僅提供 2D 影像輸入，對於本研究之 3D 影像無法提供 pretrain 模型，因此在 pretrain 資料的選擇上是使用公開的結節資料庫 LIDC-IDRI (Lung Image Database Consortium and Image Database Resource Initiative) 的資料進行 pretrain，同時也對其進行資料增量讓訓練樣本增加，而 pretrain 的目標為分類腫瘤的良惡性，希望透過分類腫瘤的良惡性讓模型先學一些紋理資訊，並提供分類是否具侵犯性腫瘤時一些侵犯性的特徵，下圖 3.39 為簡易流程圖。Pretrain 的結果在正確率、靈敏度、特異度與



AUC 的表現為：80%、66.67%、93.33%、0.8044，經過本研究之資料 finetune 過後正確率、靈敏度、特異度與 AUC 為：56.18%、64.44%、51.55%、0.6311。

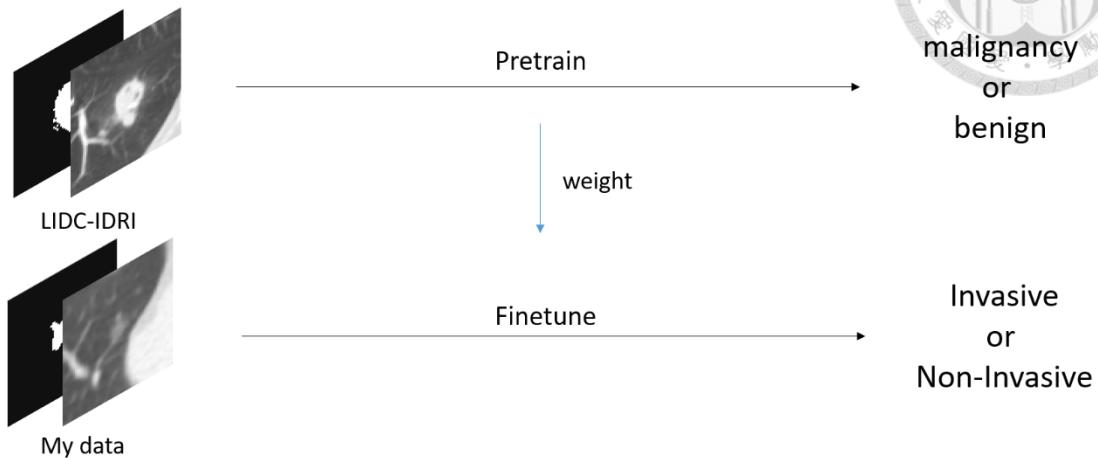
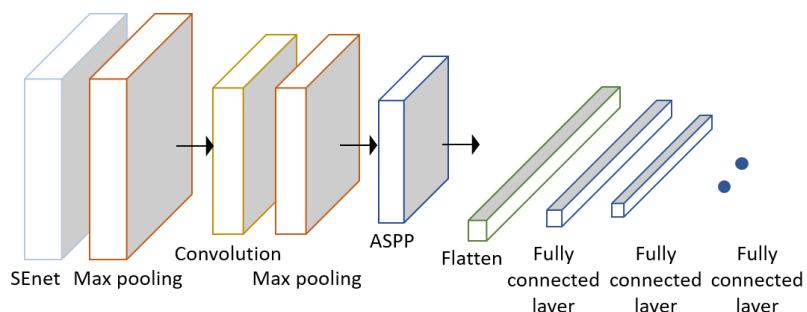


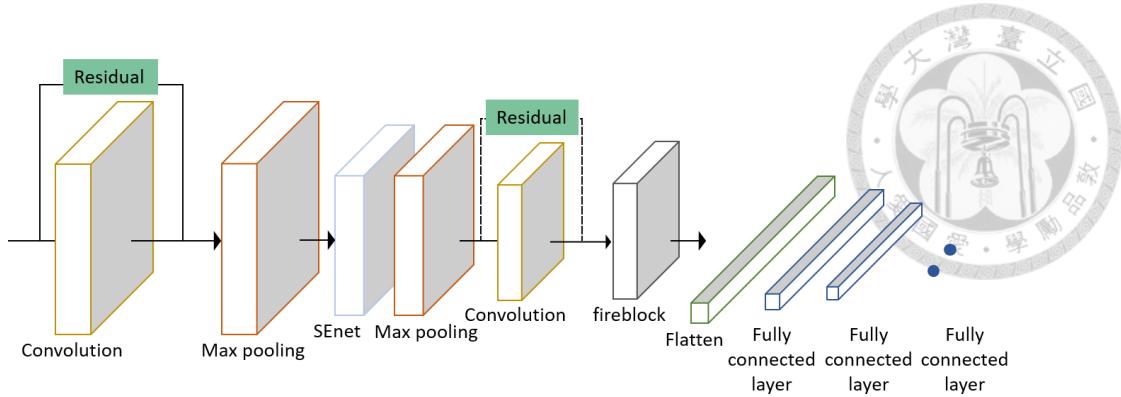
圖 3.39 pretrain 及 finetune 簡易流程圖

3.3.3.4. Ensemble learning

Ensemble learning[32]的精神在於集成多個分類器的結果相較於單一個模型的結果來的要好，當 ensemble learning 用在深度學習便可使用不同的模型架構對目標資料各自進行訓練，學習其認為重要的特徵，訓練完後再將這些模型的結果做平均或是加權平均並輸出最後的預測結果。在先前研究利用兩個模型進行訓練並平均預測結果，使用的模型如圖 3.40。



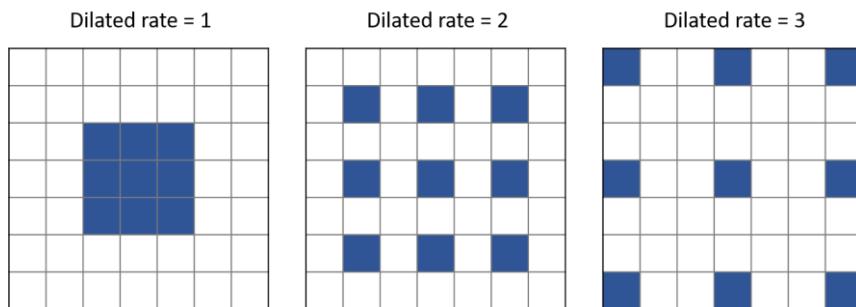
(a) Model 1



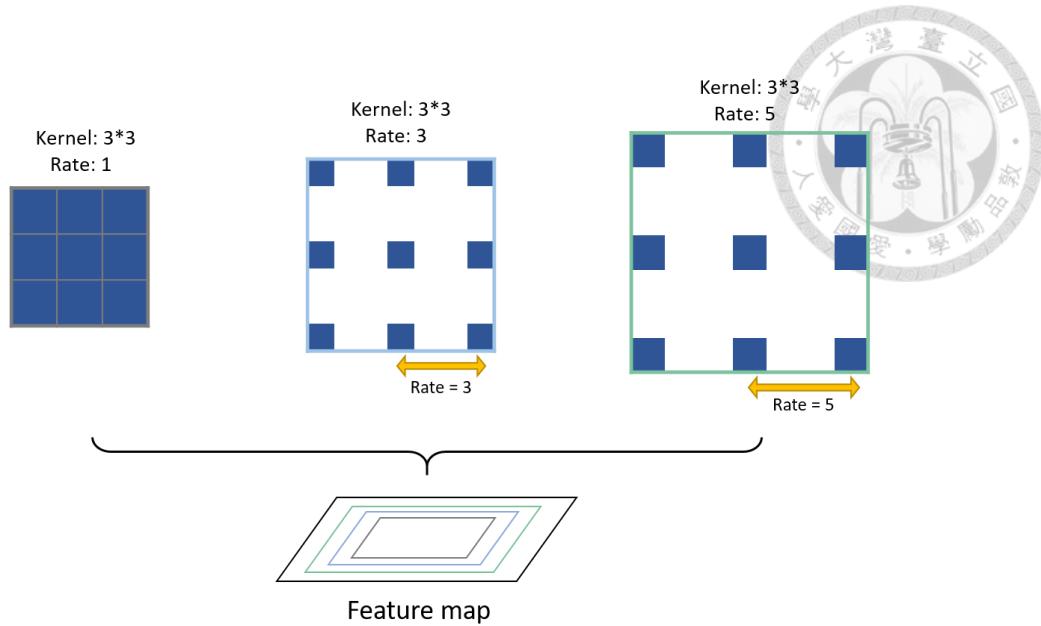
(b) Model 2

圖 3.40 Ensemble 模型架構(a) model 1 (b) model 2

於 Model 1 有使用一個子架構 ASPP(Atrous Spatial pyramid pooling)[33]，此子架構的目的是希望使用不同採樣率的空洞卷基層(Atrous convolution)，來捕捉不同尺度的訊息，同時提取小範圍與大範圍的影像資訊，而空洞卷基層為透過在標準的 convolution kernel 中輸入空洞（參數 Dilate rate 即可調整其放大的大小）使其在不增加參數量的前提下達到增加感受野的目的，圖 3.41。



(a) Atrous convolution



(b) Atrous spatial pyramid pooling

圖 3.41 (a) Atrous convolution (b) Atrous spatial pyramid pooling

另外在 Model 2 有兩個子架構分別是 residual block 與 fire block，其中 residual block 是為了解決梯度消失的問題而產生的[34]，在除了原本做 convolution + activation 的通道外，額外產生了一條捷徑通道，這個捷徑不會經過前面提及的通道，而是直接和通道的 output 進行相加再繼續後面的計算，少了線性與非線性轉換的捷徑解決了當網路加深時，gradient 不能回流下一層的問題，如圖 3.42。另外是 fire block，此子架構是來自於 squeeze net[35]，其目的為壓縮 feature map 的大小以減少參數量的增加，過程中有三個步驟：首先，先通過 1×1 的卷積減少通道數，再利用 $1 \times 1 \times 1$ 與 $3 \times 3 \times 3$ 的 kernel 進行卷積，最後將其卷積結果 concatenate 在一起，示意圖如圖 3.43。

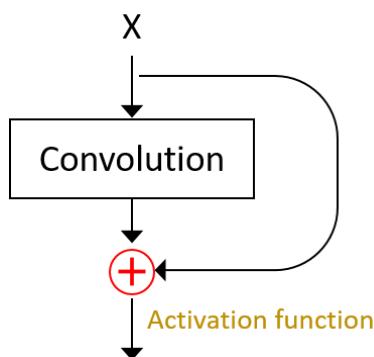


圖 3.42 Residual block

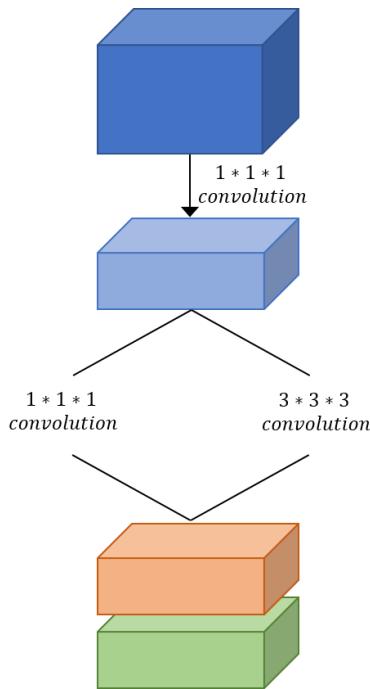


圖 3.43 fireblock

Model 1 訓練完之正確率、靈敏度、特異度與 AUC 分別為：56.18%、64.44%、51.55%、0.5902，Model 2 則是：56.97%、62.22%、54.04%、0.5969，透過 ensemble learning 的方式將這兩個模型預測結果進行平均得到之預測結果正確率、靈敏度、特異度與 AUC 為：55.78%、64.44%、50.93%、0.5831，此結果相較於兩個模型各自訓練之結果退步，在先前嘗試的 ensemble learning 皆有此現象，由此判斷可能在模型還沒有達到一定的判斷能力前，做 ensemble learning 的效果有限。

3.3.3.5. 加入 feature map 做 training

本研究在進行深度學習模型訓練前，有嘗試使用機器學習的方式尋找適合分類 IA 與 non-IA 之特徵，故此章節將挑出幾個特徵將其製成 feature map 之 VOI 同時丟入模型當中進行訓練，除了本研究先前挑選之較佳特徵外，也放入較為簡單之一階特徵以及先前實驗室對於 high grade 研究所提取之特徵，最後加上影像後共有七種不同輸入 VOI 同步訓練，下表 3.15 為所有模型輸入 VOI。

表 3.15 模型輸入之 feature map



模型輸入	
1	原始影像
2	Cluster shade
3	Entropy
4	Energy
5	Homogeneity
6	Percentile75
7	Sum of squares

此訓練方式為每個 fold 皆平行進行 7 個模型之訓練，每一個模型的 input 各自放入其中一種輸入 VOI，並且同樣結合 adaboost 之方式更改每個模型其樣本 sample weight 以及透過 sample weight 產生之區間抽取 1000 筆資料再 transfer 其各自最好的權重後，再繼續第二次模型訓練，如此重複五次（即為進行 5 回合，每回合之訓練圖為圖 3.44）才完成一個 fold 的訓練，再從各個回合中得到最好之結果來對最後每個 fold 之結果進行計算，等到五個 fold 皆結束後，也會再次透過每個 fold 之結果計算最終效能（正確率、靈敏度、特異度及 AUC），其結果如下表 3.16：



圖 3.44 每回合進行之模型訓練圖



表 3.16 原始模型與多 feature map 輸入之結果比較

	原始影像	僅腫瘤部分影像
Accuracy	0.6425 ± 0.027	0.5387 ± 0.021
Sensitivity	0.6533 ± 0.031	0.5333 ± 0.013
Specificity	0.6222 ± 0.020	0.5410 ± 0.024
AUC	0.6667 ± 0.049	0.5663 ± 0.008



第四章研究結果與討論

在 3.2.3.7 這節有介紹本研究所提倡之模型架構，本章將針對提倡之模型架構與參數進行比較，呈現不同參數或者架構對於模型的影響，另外也會與沒有對研究資料做 solid ratio 的限制做結果比較。

4.1 針對分不好的資料進行著重學習的三種方式

這邊將比對三種針對分得比較不好的資料進行著重學習的方式，第一個是透過前一次的模型分類結果來決定樣本 sample weight 的改變，根據式(3.9)計算進入下一個模型訓練之前，每一筆資料權重的調整，以著重分類效果比較不好的資料，而較好的則給予較小的權重；第二個為透過計算每筆資料進入下一個模型之前產生對應的權重後，再利用新的權重製造出區間並且隨機抽取介於 0 到 1 之間的變數，對訓練樣本資料做抽樣，總共抽取 N（本研究嘗試了 1000、1500、2000 以比較其結果）筆資料加入原本的樣本中，由於區間是由權重當作比例而劃分的，所以對於權重較大的資料區間較大進而使抽中機率的也越高，加入抽樣資料後每筆樣本的權重又會重新設定成相同的權重（也就是 $1/\text{訓練樣本數}(\text{原樣本數} + \text{抽取樣本數})$ ）；第三者則是同時進行，也就是本研究所使用之方法，計算完進入下一個模型前每筆資料的權重後，加入抽樣資料，和第二種不一樣的地方在於這個方式並不會把資料權重重新設定為相同的權重，而是使用調整過後的權重繼續做訓練。

於第一種僅針對資料權重做更改的方式重複進行三次的 5-fold cross validation 的 AUC 結果為： 0.6240 ± 0.015 ；第二種利用更改的資料權重做樣本抽取的方式其餘抽取 1000、1500、2000 筆資料進行三次 5-fold cross validation 分別的平均 AUC 為： 0.5922 ± 0.023 、 0.5958 ± 0.007 、 0.6065 ± 0.013 ，四者之 AUC 變化沒有太大的差別，而由此結果相較於第一種來說改變資料權重較能夠有效的訓練



資料，獲得較好的結果；第三種同時更改權重與抽取資料的方式，在各種參數測試下達到最好的效果，由於於第一種的方式測試下不同的抽取數量其 AUC 沒有明顯的差距，故在此方式下也測試了幾種抽取的數量分別為：1000、1500、2000，其於三次的 5-fold cross validation 訓練完之平均 AUC 分別為 0.5955 ± 0.014 、 0.6425 ± 0.027 、 0.6095 ± 0.029 ，這三者結果比較以兩種同時使用之方法有較好的判斷，故本研究採用同時使用之方法進行迭代，三者比較結果如表 4.1，表 4.2 表 4.2 為最佳結果之正確率、靈敏度與特異度。

本研究推論，由於每筆資料已透過上一次訓練結果更改過權重，加入太少的資料可能不會產生太多的幫助、太多則可能導致學習較差的資料與學習較好的資料得到過於不均的針對程度，導致結果也下降。

表 4.1 分類較不好的樣本三種著重學習的方式之 AUC 結果

加入資料數量	方法	方法一	方法二	方法三
1000			0.5922 ± 0.023	0.5955 ± 0.014
1500			0.5958 ± 0.007	0.6425 ± 0.027
2000			0.6065 ± 0.013	0.6095 ± 0.029
			0.6240 ± 0.015	

表 4.2 方法三最佳方法之正確率、靈敏度、特異度

	Accuracy	Sensitivity	Specificity
方法三 (加入 1500 筆資料)	0.6533 ± 0.031	0.6222 ± 0.020	0.6667 ± 0.049

4.2 kernel number 數量的影響

kernel number 也是影響模型訓練結果的一個變因，多數模型在設置 kernel number 時會從淺層至深層逐漸加大，讓模型學到更多深層語義資訊，但由於本研



究受限於沒有大量的樣本資料可以進行訓練，若使用大量的參數來對資料做訓練，容易導致過擬化情形的發生，所以為減少訓練參數以減少這種情況發生進而選擇較低的 kernel number 數量，本研究將透過前面測試出最佳更新權重的訓練方式的模型，針對 kernel number 設置為 4、8、16，並且同樣增加 1500 筆資料之結果對其進行比較，下表 4.3 為其結果：

表 4.3 不同 kernel number 之 AUC 結果

AUC \ Kernel number	4	8	16
AUC	0.6425 ± 0.027	0.6199 ± 0.023	0.6026 ± 0.057
Accuracy	0.6533 ± 0.031	0.5940 ± 0.020	0.5920 ± 0.028
Sensitivity	0.6222 ± 0.020	0.5934 ± 0.010	0.5867 ± 0.038
Specificity	0.6667 ± 0.049	0.5943 ± 0.024	0.5943 ± 0.024

於 kernel number 為 8 及 16 之結果來看，其訓練沒有太大的差異，AUC 分別為 0.6199 ± 0.023 、 0.6026 ± 0.057 ，但可以觀察到在 kernel number 為 4 時，其結果較佳為 0.6425 ± 0.027 ，此結果表現出降低 kernel number 的數量可以使模型達到更好的訓練效能，故將 kernel number 設置為 4，讓模型減少參數量。

4.3 Attention layer 的重要性

本研究提倡之模型其中一子架構為 attention layer，這個子架構幫助模型更著重於腫瘤內部的學習，本節將比較（1）沒有加入 attention layer 的模型與（2）有加入 attention layer 的模型做完三次 5-fold cross validation 的結果，而 attention layer 放置的位置為同時放在第一個 block 與第二個 block 結束的地方，也是本研究提出之架構，如所圖 4.1 所示。

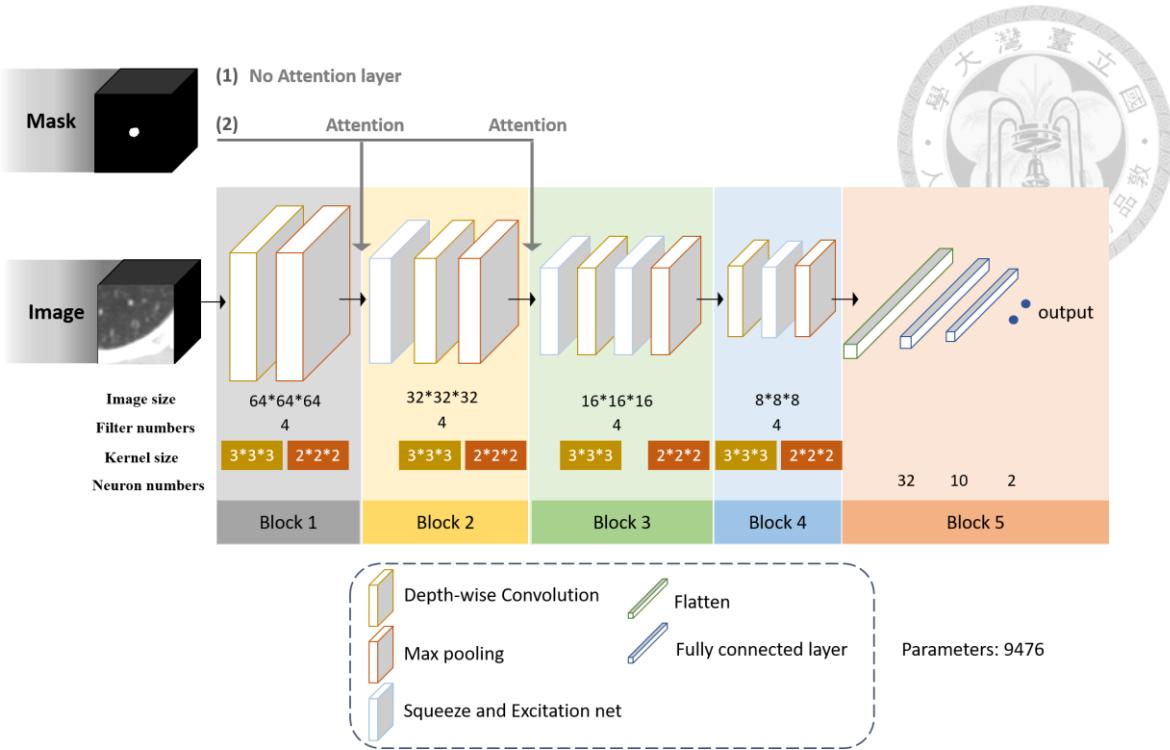


圖 4.1 Attention layer 的位置

沒有加入 attention layer 前模型學習的效果十分不好，模型得到之結果在 AUC 的結果為： 0.4983 ± 0.027 ； block 1, 2 後面都加入 attention layer 即為本章於 4.1 節之訓練結果較佳模型： 0.6425 ± 0.027 。從結果可以觀察到加入 attention layer 其結果與沒加入的提升許多，從原本平均下不到 0.5 的 AUC 提升至 0.6425，由此得知 attention layer 對於偵測毛玻璃狀腫瘤是一個重要的架構，在目標物體與背景顏色沒有明顯的對比下，此架構可以讓原本影像上較為模糊不清楚的主體在去除主體以外的干擾後變得比較好偵測以及學習，使模型更能針對腫瘤進行參數調整，以有效學習到毛玻璃狀之肺腺癌在具侵犯性與不具侵犯性兩類各自的特徵。

4.4 影像輸入

如 4.3 節，本研究利用腫瘤的 binary mask 進行 attention 來提升模型對於腫瘤內部的學習效果，所以本節將嘗試將影像的輸入直接放進只有腫瘤部分之 VOI，並將背景設置為 0，以觀察結果，其 VOI 的影像如下圖 4.2 所示：

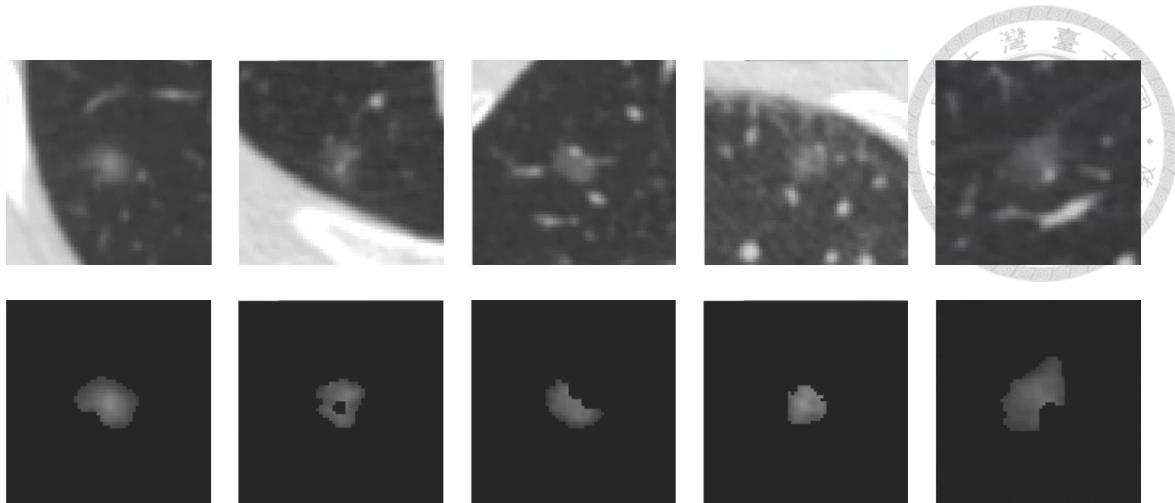


圖 4.2 原始 VOI (上排) 與只有腫瘤部分之 VOI (下排)

下表為原始 VOI 與放入僅有腫瘤部分其餘背景為 0 之 VOI 得到的結果，其中可觀察到若僅留下腫瘤部分背景為 0 之 VOI 當作輸入之結果下降，並沒有對模型的學習效能有所增加，判斷對於腫瘤周圍仍有一些重要的資訊讓模型進行學習，讓模型先透過整個完整原始影像 VOI 進行學習後，對於腫瘤邊界仍可學習到一些背景部分，再予以 binary mask 對 feature map 進行 attention，才有較好的結果。

表 4.4 原始 VOI 與僅有腫瘤部分之 VOI 效果比較圖

	原始影像	僅腫瘤部分影像
Accuracy	0.6425 ± 0.027	0.5387 ± 0.021
Sensitivity	0.6533 ± 0.031	0.5333 ± 0.013
Specificity	0.6222 ± 0.020	0.5410 ± 0.024
AUC	0.6667 ± 0.049	0.5663 ± 0.008

4.5 fusion

於 2.3 節有提到 Xianwu Xia[13]等人所撰寫的一篇文獻是利用 deep learning model 與 radiomics model 取兩模型預測結果進行融合當作最後的預測值，而在 3.3.2 節曾使用本研究資料重複文獻作法，用其模型進行 deep learning model 的訓練，但其訓練效果相對於本研究之訓練效果差一點，所以本節會利用本研究資料與提倡之模型架構與 radiomics model 進行融合，並且同樣對融合的方式進行修



正，radiomics 會分成 training, validation, testing 並透過 validation data 融合兩模型取得 AUC 最高的結合方式再套用至 testing data 觀察結果。在深度學習模型訓練結果如 4.1 所示，AUC 為 0.6425 ± 0.027 ；另外單純以 radiomics 進行建模得到的測試集結果如 3.3.2 所示為：55.20%、58.67%、53.71%、0.5566。

當輸入影像為原本的 VOI (仍保留背景資訊)，取其中一次的訓練的模型之 validation data 對 deep learning model 與 radimoics model 進行融合，得到最佳 AUC 的結合方式為：深度模型預測值 * 0.8 + radiomics 模型預測值 * 0.2，AUC 為 0.7092，接著從 validation 挑出最好的結合方式套用至 testing 得到的正確率、靈敏度、特異度與 AUC 分別為：60.80%、62.67%、60.00%、0.6384，其 AUC 與深度學習所得到的結果十分相近。三者 AUC 比較如圖 4.3：

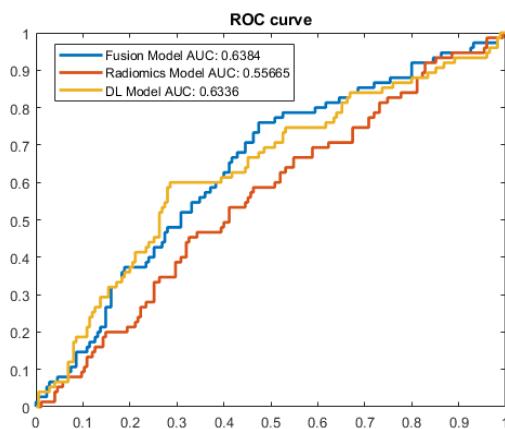


圖 4.3 輸入影像為原始影像之深度、radiomics、fusion 模型 ROC 曲線

雖然不同於原本文獻是直接使用 testing data 進行融合，本研究將 radiomics 拆分成 training, validation 與 testing 後其評估效能減弱，但對於做 fusion 模型仍有一些參考價值，使的最後 AUC 仍有些微提。

4.6 Solid ratio 的影響

本章節除了會針對（1）本研究資料 solid ratio <0.25 所得到之結果（histogram 分佈如圖 4.4 (a)）與（2）資料包含 solid ratio > 0.25 之結果（histogram 分佈如圖 4.4 (b)）進行比較外，為避免是資料數量增加而造成（2）之結果提升，故另外會



從（3）包含 solid ratio > 0.25 的所有資料中挑選與（1）相同數量之 IA、Non-IA，其中 IA 選擇 HU 值分佈相對較高的、Non-IA 則選擇 HU 值相對低的資料（histogram 分佈如圖 4.4 (c)），一起進行結果比較。在本研究使用 solid ratio < 0.25 的資料當中，Invasive Adenocarcinoma 平均 CT value 為 -704.3 ± 48.3 ($-860.1, -604.2$)，Non-Invasive Adenocarcinoma 平均 CT value 為 -726.8 ± 50.1 ($-826.0, -606.7$)；而包含 solid ratio > 0.25 之資料其 IA 平均 CT value 為 -649.9 ± 79.1 ($-843.4, -422.0$)， Non-IA 平均 CT value 為 -701.1 ± 61.4 ($-818.6, -542.9$)；另外，在極端的資料其 IA 平均 CT value 為 -595.9 ± 61.4 ($-664.2, -422.0$)， Non-IA 平均 CT value 為 -719.3 ± 46.7 ($-818.6, -638.0$)。（1）與（2）之 HU 比較，不論在 IA 或是 non-IA 其平均值皆有提升，但由於本研究仍挑選呈現較 GGO 之腫瘤，所以平均 CT value 仍落在 -400 以下，三者之 HU 統計見表 4.5、三者之 IA 與 Non-IA 個數統計見表 4.6。

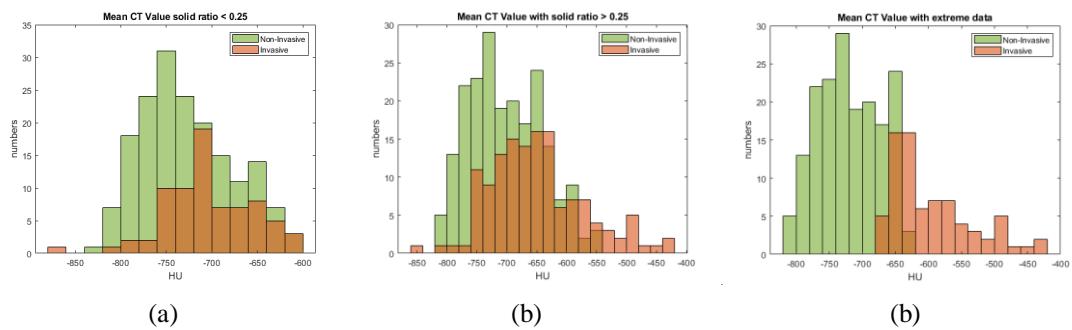


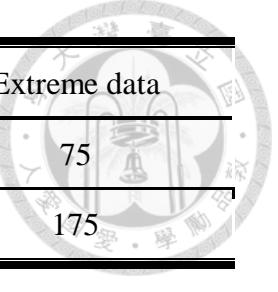
圖 4.4 mean CT value histogram

表 4.5 三種資料之 mean CT value 統整表

Limitation	Solid ratio < 0.25	No limitation	Extreme data
IA	-704.3 ± 48.3 ($-860.1, -604.2$)	-649.9 ± 79.1 ($-843.4, -422.0$)	-595.9 ± 61.4 ($-664.2, -422.0$)
Non-IA	-726.8 ± 50.1 ($-826.0, -606.7$)	-701.1 ± 61.4 ($-818.6, -542.9$)	-719.3 ± 46.7 ($-818.6, -638.0$)

*括號內為(最小 HU 值,最大 HU 值)

表 4.6 三種資料之 IA、NonIA 個數比較



	Solid ratio < 0.25	No limitation	Extreme data
IA	75	136	75
Non-IA	175	207	175

限制於 solid ratio < 0.25 的資料訓練相對較佳的模型在輸入端為原始影像 VOI 其 AUC 為： 0.6425 ± 0.027 ，於是將對加入沒有限制 solid ratio 的資料以相同的輸入與模型進行訓練，而由於加入沒有限制 solid ratio 的資料後，兩類別資料比例有些微調整，所以在調整參數過後，得到的最佳結果為： 0.6738 ± 0.014 。雖在 IA 與 NonIA 加入不限制 solid ratio 的資料後在 mean CT value 這項指標的數值有些微提升，變化幅度不至於非常大，但已可以讓模型訓練的效果有所提升，相較於前面文獻回顧提供之 mean CT value 兩類差距都較大，可見 mean CT value 對於此類別之判斷是有一定程度的影響力；再者，挑選出與具有 solid ratio 限制的相同數量的兩類別極端資料其平均 AUC 為： 0.9164 ± 0.016 ，可見於第二組之效果提升並非是因為其資料量的增加而有所上升，HU 值的差別確實對於本實驗的分類是有幫助的。

表 4.7 三種資料之平均 AUC

	Solid ratio < 0.25	No limitation	Extreme data
AUC	0.6425 ± 0.027	0.6738 ± 0.014	0.9164 ± 0.016
Accuracy	0.6533 ± 0.031	0.6501 ± 0.013	0.8347 ± 0.015
Sensitivity	0.6222 ± 0.020	0.6569 ± 0.023	0.8267 ± 0.013
Specificity	0.6667 ± 0.049	0.6457 ± 0.035	0.8381 ± 0.017



第五章結論與未來展望

5.1 結論與未來展望

根據衛生福利部於 109 年國人死因統計結果[1]，氣管、支氣管、肺癌位於癌症死因中的第一且已蟬聯多年，是一項不可輕忽的疾病，若能及早給予患者相應的治療的話，可以相對提高存活率。而較早期的肺部腫瘤在電腦斷層掃描上通常會呈現毛玻璃狀，針對這些毛玻璃狀腫瘤做是否具侵犯性的區分，對於醫師評估是否需進行手術切除及其預後給予參考資訊。

鑑於 solid ratio 以及最大直徑皆為對於分類是否具侵犯性的腫瘤一個很重要的判斷依據，本研究為針對更難以區分之毛玻璃狀腫瘤對這兩項變因做出限制，收集 solid ratio < 0.25 且最大直徑<3cm 之毛玻璃狀影像，透過建立深度學習模型的方式，針對這些資料進行是否具侵犯性之分類。本研究以 AdaBoost 為研究主軸，目的是希望可以對於模型分類較不好的影像著重學習，即根據前一次模型訓練結果調整每筆訓練資料的權重，並同時透過新的權重建立每筆樣本位於 0 到 1 間的區間，再產生一個隨機的小數，對訓練資料做 1500 次重複抽樣，當權重越大表示其抽中的機率越大，增加其加入訓練樣本的機會，讓模型多做學習。在 kernel number 的選擇上，會選擇較小的數量，原因是考量到本研究收集之資料數量不足，無法透過大量的參數做訓練，為避免導致過擬化的現象所做的選擇，在測試結果中也可以得知並非 kernel number 數量越多就可以得到越好的結果。另外，對於 attention layer 對於分類效果的影響相當顯著，從沒有加入 attention layer 與加入 attention 的結果其 AUC 可以從 0.4983 提升至 0.6425，提升的幅度很大，對於模型之訓練有很大的幫助。最後 solid ratio 對於分類的影響，本研究加入呈現毛玻璃狀，但 solid ratio 大於 0.25 之腫瘤進行訓練，根據結果可以觀察到其 AUC 也有相當程度的提升（從 0.6425 提升至 0.6738），若提取更極端之資料進行

訓練，其 AUC 甚至從 0.6425 攀升至 0.9164，可見 solid ratio 為一重要判斷因子。

本研究由於資料數量不足，對於 data driven 的深度學習來說，如何增加資料量是一大難題，除了本研究使用的旋轉位移等方式來做資料增量外，是否有機會找到一個更好、可以找到像是 2D 的 imagenet 一樣對 3D 影像做 pretrain 的模型，可能為一個可以考慮的面向。

5.2 研究限制

- (1) 本研究目前所收集之資料僅有 250 筆，對於做深度學習來說資料量有些不足，若可以增加更多的資料，應該會對訓練結果有幫助。
- (2) 由於本研究針對的是 solid ratio 小於 0.25 之腫瘤，而腫瘤具有多一點 solid part 的比例在分類上多為屬於 invasive adenocarcinoma，所以相較下對於 solid ratio 有所限制會提升其分類難度；另外腫瘤之大小的限制也是會有影響，對於較大的腫瘤在分類上多為屬於 invasive adenocarcinoma，同樣會提升分類難度。
- (3) 本研究難以從有限數量之樣本資料找到適合區分兩類的 radiomics 的特徵，若可以找到更適合的特徵，在使用 fusion 的方法上也可以提升其訓練效能。



參考資料

- [1] 衛生福利部 109 年國人死因統計結果
- [2] Travis, W. D., Brambilla, E., Noguchi, M., Nicholson, A. G., Geisinger, K. R., Yatabe, Y., Beer, D. G., Powell, C. A., Riely, G. J., & Van Schil, P. E. (2011). International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma. *Journal of Thoracic Oncology*, 6(2), 244-285.
- [3] Liu, S., Wang, R., Zhang, Y., Li, Y., Cheng, C., Pan, Y., Xiang, J., Zhang, Y., Chen, H., & Sun, Y. (2016). Precise diagnosis of intraoperative frozen section is an effective method to guide resection strategy for peripheral small-sized lung adenocarcinoma. *Journal of clinical oncology*, 34(4), 307-313.
- [4] Nesbitt, J. C., Putnam Jr, J. B., Walsh, G. L., Roth, J. A., & Mountain, C. F. (1995). Survival in early-stage non-small cell lung cancer. *The Annals of thoracic surgery*, 60(2), 466-472.
- [5] Jemal, A., Siegel, R., Xu, J., & Ward, E. (2010). Cancer statistics, 2010. *CA: a cancer journal for clinicians*, 60(5), 277-300.
- [6] Detterbeck, F. C. (2018). The eighth edition TNM stage classification for lung cancer: What does it mean on main street? *The Journal of thoracic and cardiovascular surgery*, 155(1), 356-359.
- [7] Travis, W. D., Asamura, H., Bankier, A. A., Beasley, M. B., Detterbeck, F., Flieder, D. B., Goo, J. M., MacMahon, H., Naidich, D., & Nicholson, A. G. (2016). The IASLC lung cancer staging project: proposals for coding T categories for subsolid nodules and assessment of tumor size in part-solid tumors in the forthcoming eighth edition of the TNM classification of lung cancer. *Journal of*

Thoracic Oncology, 11(8), 1204-1223.

- 
- [8] Fan, L., Fang, M., Li, Z., Tu, W., Wang, S., Chen, W., Tian, J., Dong, D., & Liu, S. (2019). Radiomics signature: a biomarker for the preoperative discrimination of lung invasive adenocarcinoma manifesting as a ground-glass nodule. *European radiology*, 29(2), 889-897.
 - [9] Cho, H.-h., Lee, G., Lee, H. Y., & Park, H. (2020). Marginal radiomics features as imaging biomarkers for pathological invasion in lung adenocarcinoma. *European radiology*, 30(5), 2984-2994.
 - [10] Wu, L., Gao, C., Xiang, P., Zheng, S., Pang, P., & Xu, M. (2020). CT-imaging based analysis of invasive lung adenocarcinoma presenting as ground glass nodules using peri-and intra-nodular radiomic features. *Frontiers in oncology*, 10, 838.
 - [11] Wu, G., Woodruff, H. C., Shen, J., Refaee, T., Sanduleanu, S., Ibrahim, A., Leijenaar, R. T., Wang, R., Xiong, J., & Bian, J. (2020). Diagnosis of invasive lung adenocarcinoma based on chest CT radiomic features of part-solid pulmonary nodules: a multicenter study. *Radiology*, 297(2), 451-458.
 - [12] Gong, J., Liu, J., Hao, W., Nie, S., Zheng, B., Wang, S., & Peng, W. (2020). A deep residual learning network for predicting lung adenocarcinoma manifesting as ground-glass nodule on CT images. *European radiology*, 30(4), 1847-1855.
 - [13] Xia, X., Gong, J., Hao, W., Yang, T., Lin, Y., Wang, S., & Peng, W. (2020). Comparison and fusion of deep learning and radiomics features of ground-glass nodules to predict the invasiveness risk of stage-I lung adenocarcinomas in CT scan. *Frontiers in oncology*, 418.
 - [14] Suzuki, K., Koike, T., Asakawa, T., Kusumoto, M., Asamura, H., Nagai, K., Tada, H., Mitsudomi, T., Tsuboi, M., & Shibata, T. (2011). A prospective radiological



- study of thin-section computed tomography to predict pathological noninvasiveness in peripheral clinical IA lung cancer (Japan Clinical Oncology Group 0201). *Journal of Thoracic Oncology*, 6(4), 751-756.
- [15] Suzuki, K., Watanabe, S.-i., Wakabayashi, M., Saji, H., Aokage, K., Moriya, Y., Yoshino, I., Tsuboi, M., Nakamura, S., & Nakamura, K. (2022). A single-arm study of sublobar resection for ground-glass opacity dominant peripheral lung cancer. *The Journal of thoracic and cardiovascular surgery*, 163(1), 289-301. e282.
- [16] Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1), 119-139.
- [17] Duda, R. O., Hart, P. E., & Stork, D. G. (2000). Pattern Classification. John Wiley & Sons. Inc., New York, 5.
- [18] Ye, R., Liu, F., & Zhang, L. (2019). 3D depthwise convolution: reducing model parameters in 3D vision tasks. Canadian Conference on Artificial Intelligence,
- [19] Lin, M., Chen, Q., & Yan, S. (2013). Network in network. *arXiv preprint arXiv:1312.4400*.
- [20] Hertz, J., Krogh, A., & Palmer, R. G. (2018). *Introduction to the theory of neural computation*. CRC Press.
- [21] Glorot, X., Bordes, A., & Bengio, Y. (2011). Deep sparse rectifier neural networks. Proceedings of the fourteenth international conference on artificial intelligence and statistics,
- [22] Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. Proceedings of the IEEE conference on computer vision and pattern recognition,
- [23] Taherkhani, A., Cosma, G., & McGinnity, T. M. (2020). AdaBoost-CNN: An



- adaptive boosting algorithm for convolutional neural networks to classify multi-class imbalanced datasets using transfer learning. *Neurocomputing*, 404, 351-366.
- [24] Hastie, T., Rosset, S., Zhu, J., & Zou, H. (2009). Multi-class adaboost. *Statistics and its Interface*, 2(3), 349-360.
- [25] Lim, H.-j., Ahn, S., Lee, K. S., Han, J., Shim, Y. M., Woo, S., Kim, J.-H., Yie, M., Lee, H. Y., & Chin, A. Y. (2013). Persistent pure ground-glass opacity lung nodules \geq 10 mm in diameter at CT scan: histopathologic comparisons and prognostic implications. *Chest*, 144(4), 1291-1299.
- [26] Grove, O., Berglund, A. E., Schabath, M. B., Aerts, H. J., Dekker, A., Wang, H., Velazquez, E. R., Lambin, P., Gu, Y., & Balagurunathan, Y. (2015). Quantitative computed tomographic descriptors associate tumor shape complexity and intratumor heterogeneity with prognosis in lung adenocarcinoma. *PloS one*, 10(3), e0118261.
- [27] Son, J. Y., Lee, H. Y., Lee, K. S., Kim, J.-H., Han, J., Jeong, J. Y., Kwon, O. J., & Shim, Y. M. (2014). Quantitative CT analysis of pulmonary ground-glass opacity nodules for the distinction of invasive adenocarcinoma from pre-invasive or minimally invasive adenocarcinoma. *PloS one*, 9(8), e104066.
- [28] Zhang, Y., Heuvelmans, M., Zhang, H., Oudkerk, M., Zhang, G., & Xie, X. (2018). Changes in quantitative CT image features of ground-glass nodules in differentiating invasive pulmonary adenocarcinoma from benign and in situ lesions: histopathological comparisons. *Clinical radiology*, 73(5), 504-509. e509-504. e516.
- [29] Chae, H.-D., Park, C. M., Park, S. J., Lee, S. M., Kim, K. G., & Goo, J. M. (2014). Computerized texture analysis of persistent part-solid ground-glass nodules: differentiation of preinvasive lesions from invasive pulmonary adenocarcinomas.

Radiology, 273(1), 285-293.

- 
- [30] Mitchell, M. (1998). *An introduction to genetic algorithms*. MIT press.
 - [31] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.
 - [32] Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4), e1249.
 - [33] Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4), 834-848.
 - [34] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition,
 - [35] Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., & Keutzer, K. (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size. *arXiv preprint arXiv:1602.07360*.