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非小細胞癌 PD-L1 表現預測：結合遮蓋圖像模型與生成對抗網路

Prediction of PD-L1 Expression in Non-Small Cell Lung Cancer on Chest CT Scans: A masked image model approach combined with a GAN method

周姵妤

Pei-Yu Chou

指導教授：陳中明 博士

Advisor: Chung-Ming Chen, Ph.D

中華民國 112 年 7 月

July, 2023

摘要

肺癌是全球常見且致死率最高的癌症，儘管醫學科學在肺癌的治療方面取得了一些重大進展，晚期肺癌的五年存活率仍然很低。近年來，針對晚期肺癌的治療中引入了標靶治療和免疫療法，這些新的治療方法為患者帶來了一些希望。目前在非小細胞肺癌治療中應用最廣的是 PD-1/PD-L1 抑制劑，透過反制腫瘤的逃脫機制，利用自身免疫反應排除腫瘤。使用 PD-1/PD-L1 抑制劑在晚期治療上的顯著效果，但仍然需要識別可以獲益於此治療的病患族群。然而，目前用於判定免疫療法適用性的 PD-L1 表現量檢測方法存在問題（例如腫瘤異質性、染色標準不一），其準確度尚有進步空間。

因此，本研究希望利用非侵入性且能夠整體判讀腫瘤的 CT 影像來建立電腦輔助診斷（CAD），進而幫助 PD-L1 表現量的檢測。目前建立 CAD 系統的方法可分為機器學習以及深度學習，深度學習模型在訓練過程中可自行提取特徵，但需要大量的標記樣本，而醫學影像樣本相對不足。近年來，自監督學習提出新的訓練架構，可以使用未標記的數據進行訓練，降低樣本門檻。縱使目前基於自監督學習的遮蓋圖像模型在自然影像尚有著不錯的成績，面對醫學影像這種目標較不明確的資料仍存在一些限制。

為了克服上述醫學影像不足與目標物不明確的問題，進而提高 PD-L1 表達的預測準確性，本研究針對醫學影像的特性提出了一 Multi-task Masked Autoencoder（MTMAE）方法。MTMAE 具有以下三個特點：（1）使用基於自監督學習的遮蓋圖像模型，使模型具有較高的遷移能力；（2）在多任務學習中加入分割任務，使模型在提取特徵時能夠區分前景和背景，更好地捕捉腫瘤的特徵；（3）使用生成對抗

網絡（GAN）生成影像，使模型能夠學習到大量多樣的特徵，以克服學習受限的問題。

通過實驗驗證，在本研究共 188 個 PD-L1 樣本上使用上述提出的模型進行 PD-L1 50%表現量分類，AUC 為 0.735，準確率為 0.724。相比於傳統的監督式預訓練（AUC : 0.695)和訓練單一重建任務的 MAE(AUC : 0.712)，本研究提出的 MTMAE 模型在實驗中表現更好。本研究結合了自監督學習、多任務學習和 GAN 生成影像的特點，針對醫學影像特性與資料量依賴性進行改善，進而應用於幫助分類 PD-L1 表現量。

關鍵字：自監督學習、遮蓋圖像模型、生成對抗網路、PD-L1 表現量、免疫治療、非小細胞肺癌、深度學習

Abstract

Lung cancer is a common and highly fatal malignancy worldwide. Despite significant advancements in medical science, the five-year survival rate for advanced-stage lung cancer remains low. Recent therapeutic strategies, such as targeted therapy and immunotherapy, have brought hope to patients with advanced lung cancer. Immunotherapy, specifically the use of PD-1/PD-L1 inhibitors, has shown promising results in late-stage treatments by counteracting the tumor's evasion mechanisms and harnessing the body's immune response to eliminate the tumor. However, accurately identifying patients who are likely to benefit from this therapy remains challenging. The current methods used to assess PD-L1 expression levels, which are crucial in determining the suitability of immunotherapy, suffer from issues such as tumor heterogeneity and inconsistent staining standards, leading to suboptimal accuracy.

To address these limitations and improve the accuracy of predicting PD-L1 expression levels, this study proposes a computer-aided diagnosis (CAD) system that utilizes non-invasive CT imaging to comprehensively analyze tumors. CAD systems can be broadly categorized into machine learning and deep learning approaches. While deep learning models have the advantage of automatically extracting features during training,

they require a substantial amount of annotated data, which is often lacking in medical imaging. In recent years, Self-Supervised Learning has introduced a new training framework that can leverage unlabeled data, thus reducing the dependency on annotated samples. However, applying existing Self-Supervised Learning-based masked image models to medical imaging, which involves less-defined targets, poses certain limitations.

To overcome the challenges associated with limited medical imaging data and unclear target objects, and to enhance the prediction accuracy of PD-L1 expression levels, this study introduces the Multi-task Masked Autoencoder (MTMAE) method, specifically tailored to the characteristics of medical imaging. The MTMAE method incorporates the following three key features: (1) harnessing a Self-Supervised Learning-based masked image model with enhanced capability in transfer learning, (2) the inclusion of a segmentation task in the multi-task learning framework to better distinguish foreground and background and capture tumor features effectively, and (3) the integration of a generative adversarial network (GAN) to generate diverse images, enabling the model to overcome learning constraints by learning a wide range of features.

Experimental validation on a dataset comprising 188 PD-L1 samples demonstrates the effectiveness of the proposed model in classifying PD-L1 expression levels using a

50% threshold, achieving an AUC of 0.735 and an accuracy of 0.724. Compared to traditional supervised pretraining (AUC: 0.695) and single-task reconstruction-based MAE (AUC: 0.712), the MTMAE model exhibits superior performance in the experiments. By combining the characteristics of Self-Supervised Learning, multi-task learning, and GAN-generated images, this study aims to address the challenges associated with medical imaging characteristics and data dependency, ultimately assisting in the accurate classification of PD-L1 expression levels.

Keywords: Self-Supervised Learning, Masked Image Modeling, Generative adversarial network, PD-L1 expression levels, Immunotherapy, Non-small cell lung cancer, Deep learning

目錄

圖目錄

表目錄

第一章 緒論

1.1 研究背景

根據全球癌症統計，肺癌是最常見的癌症，同時也是癌症主要死亡原因之一[1]。肺癌主要分為兩種，非小細胞肺癌（Non-small-cell lung carcinoma，NSCLC）與小細胞肺癌（Small-cell lung carcinoma， SCLC），多數是非小細胞肺癌， 占比為 85%。由於肺癌早期症狀不明確，約七成的肺癌患者在診斷時已處在晚期，五年存活率依擴散程度介於 6~33%之間[2]。

以往針對晚期非小細胞肺癌的治療較受限，由於腫瘤過大或擴散、轉移，不易使用手術切除，因此主要透過化學治療搭配放射線治療。然而，以化學藥劑抑制腫瘤生長缺乏專一性，正常細胞亦容易受到傷害，治療常伴隨嚴重副作用，患者在接受治療時需要同時承受病症與治療的負擔。另外，由於需要考量患者對副作用的耐受度來權衡治療強度，過往的化學藥物治療效果有限，僅能達到 5%以下的 5 年存活率[3]。

為改善過往缺乏專一性的治療，近年來 NSCLC 的治療研究也更專注在如何針對腫瘤細胞進行控制、有效減緩病情。隨著研究發展，更精準的治療策略出現，例如依據致病基因而生的標靶藥物。在肺腺癌（adenocarcinoma，ADCs）患者中以上皮生長因子受體（epidermal growth factor receptor , EGFR）基因最為常見，這個致病基因對癌細胞的分化生長與侵犯性有高度影響力。標靶治療主要藉由藥物來抑制突變腫瘤基因，並針對突變基因來擊殺癌細胞。此具有基因專一性的機制使的標靶治療的副作用相較於化療來的小，並且能夠有效改善患者 progression-free survival（PFS）的時間長度。

儘管標靶藥物的治療效果顯著，然而目前標靶藥物使用對象侷限於有 EGFR 基

因變異的肺腺癌患者，對於沒有 EGFR 基因變異或不是腺癌亞型的 NSCLC 患者則無法使用。這個難題在免疫治療的研究上發現了免疫檢查點抑制劑（Immune Checkpoint Inhibitors，ICIs）後，有了重大突破。目前在 NSCLC 治療最廣為應用的是 PD-1/PD-L1 抑制劑。腫瘤細胞表面藉由大量表現 programmed death ligand 1 (PD-L1)，來和 T 細胞的 programmed cell death protein 1(PD-1)結合，進而抑制 T 細胞的活化，以避免受到攻擊。ICIs 透過阻隔腫瘤 PD-L1 與 PD-1 的結合，反制腫瘤的逃脫機制，協助活化體內的免疫細胞，利用病患自身免疫反應排除異常的腫瘤細胞。採用 ICIs 藥物進行NSCLC 的治療能有效改善 PFS 以及 overall survival（OS）[4], [5]。

使用 PD-1/PD-L1 抑制劑在晚期 NSCLC 治療上的顯著效果，但仍然需要識別可以獲益於此治療的病患族群。當今廣泛採用的生物標誌（biomarker）為腫瘤的 PD-L1 表現程度（PD-L1 expression），並且以之作為選擇免疫治療策略的依據。然而，目前檢測 PD-L1 expression 的方法存在漏洞，其準確性仍有進步的空間。為了解決目前對於 PD-L1 表達判定的問題，並協助醫生能夠更準確地做出治療策略的判斷，我們利用電腦輔助診斷（Computer aided diagnosis, CAD）對 PD-L1 表達進行分類。CAD 的判斷結果可以作為當前 PD-L1 表達檢測的補充參考，為醫生在制定治療決策時提供輔助。對於 PD-L1 表達的分類，CAD 可以通過從大量的影像中提取、分析 PD-L1 表達特徵，進一步建立模式提供分類結果。

在開發 CAD 系統中醫學影像常面臨到數據量不足的問題，尤其是使用深度學習演算法。相比於其他領域，取得醫學影像是一項困難的任務。首先，醫學影像的標記需要專業人員完成，且需要長時間累積，成本十分高昂。其次，考慮到隱私和倫理問題，大多數醫學影像數據並不公開。由於數據的有限性，訓練模型在醫學影像分析方面變得更加具有挑戰性，可能導致模型的泛化能力不佳，難以處理未見過的情況。

近年來，自監督學習（Self-Supervised Learning）成功地解決了大型模型面臨的數據限制問題。自監督學習利用未標記的資料自行生成標籤（label）來進行訓練，不需依賴手動標記的標籤。自監督學習可以從未標記資料中學習有用的特徵，並在後續的任務中進行遷移學習（Transfer Learning）。如此，可以利用大量的未標記資料來訓練，即使只有少量標記資料可用，模型也可以獲得好的表現。著名的自然語言模型 BERT[6]，就是採用遮蓋預測任務來完成自監督訓練。BERT 遮蓋了句子中的部分字詞，使模型能夠預測並最小化所獲得預測結果與真實字詞之間的差異，通過重構句子使模型學習資料中的潛在特徵。在影像方面，同理上述的 BERT，遮蓋圖像模型（Masked Image Modeling）遮蓋部分圖塊，並從重建圖像中學習影像特徵。這種自監督預學習模型可以利用大量且無標記的影像進行訓練，再遷移至少量的標記資料做下游任務的訓練。此方法對醫學影像領域而言，大幅降低蒐集資料的難度，減少對標記影像的依賴。

1.2 PD-L1 介紹

T 細胞的主要功能是抗原專一性的細胞毒殺作用與引導其他免疫反應， T 细胞的活化主要由兩種路徑來控制。一是 T 細胞抗原受體（T Cell Receptor, TCR）識別並結合抗原呈現細胞（Antigen-Presenting Cell, APC）表達的主要組織相容性複合體（Major Histocompatibility Complex，MHC），使 T 細胞受到刺激。二是 T 細胞激活後，和 APC 相互作用的分子信號傳遞，這些增強信號分子與抑制信號分子，決定了 T 細胞的免疫反應強弱。這個利用 T 細胞表面受體以及 APC 表現的配體結合相互作用的調控機制，稱為免疫檢查點。正常情況下，T 細胞會活化來對抗外來細胞或病原。免疫檢查點作為免疫系統的保護機制，目的是為了防止 T 細胞過度反應，攻擊正常細胞。然而，本應該被辨識的癌細胞卻從此檢查機制逃逸，當癌細胞表現一到多個抑制性免疫檢查點時，將使得自身免疫細胞無法活化，不能攻擊、對抗癌細胞。觀察到這個現象後，免疫治療的研究方向開始導向阻斷腫瘤免疫檢查

點的抑制路徑。

1992 年，首先在小鼠的 T 細胞上發現 PD-1 蛋白[7]，並發現其有抑制免疫反應的功能[8]。接著，PD-1 蛋白的配體 PD-L1[9]被找到，並且 PD-L1 與結合 PD-1 會抑制 T 細胞的增生、活化以及分泌細胞因子。隨著 PD-1 與 PD-L1 的相關研究推進，人體內的腫瘤被觀察到會大量表現 PD-L1，並以此機制躲避免疫反應[10]。在發現 PD-1/PD-L1 通道對 T 細胞的抑制調控與腫瘤藉此表現 PD-L1 來逃脫免疫檢查點後，針對阻斷 PD-1 和 PD-L1 結合的免疫藥物開始開發，也就是 PD-1/PD-L1 抑制劑。PD-1/PD-L1 抑制劑透過阻隔腫瘤 PD-L1 與 PD-1 的結合，來避免兩者相互作用抑制 T 細胞的信號傳遞，幫助免疫細胞得以進行常規免疫程序，進入腫瘤內產生免疫反應。

阻擋 PD-1/PD-L1 結合的抗體藥物分為兩類，一種是目標為限制 PD-1 蛋白的 anti PD-1 藥物，代表性的為 pembrolizumab 和 nivolumab。另一種則是限制癌細胞表面配體 PD-L1 蛋白的 anti PD-L1 藥物，例如 Atezolizumab。anti PD-1 藥物作用在細胞上，正常細胞的 PD-L1 蛋白也受到抑制，有機會被 T 細胞誤傷。相較之下，anti PD-L1 藥物作用於癌細胞，安全性較高，副作用也較少。相反的，因為 anti PD-1 藥物作用效果較大，觀察到的反應率稍微高過 anti PD-L1[11]。概括來看，兩者差異並不大，在晚期 NSCLC 的病患身上也都證實了治療效果優過於傳統的化療，副作用與毒性也相較的低，已經是晚期 NSCLC 重要的治療角色。

雖然使用 PD-1/PD-L1 抑制劑已顯示了在晚期 NSCLC 治療上的顯著效果，但也並非是適合每一位病患的標準治療答案。由於 PD-1/PD-L1 抑制劑是基於對 PD-1/PD-L1 的抑制來腫瘤進行免疫反應，所以在治療前檢測腫瘤的 PD-L1 表現程度有助於篩選適合進行 PD-1/PD-L1 抑制劑藥物治療的患者。研究指出腫瘤的 PD-L1 expression 的確與 PD-1/PD-L1 抑制劑治療的效果有正相關[4], [12]。因此，腫瘤的

PD-L1 expression 已成為當今廣泛採用的 biomarker。

目前檢測 PD-L1 expression 的方式為免疫組織化學染色(Immunohistochemistry, IHC)，IHC 是一種利用抗體與抗原結合來檢測細胞特定蛋白表現量的常見方法，由螢光訊號來表示偵測到的目標抗體與檢測蛋白表現量。PD-L1 expression 的高低以 PD-L1 的腫瘤比率分數（Tumor Proportion Score, TPS）來定義，也就是對腫瘤檢體做 IHC 抗體染色檢查，PD-L1 expression 的數值即為檢體腫瘤細胞數分之檢體中染色陽性的腫瘤細胞數。通常以兩種截斷值為主，一是 TPS 50%，以 ≥ 50%代表高 PD-L1 表達，反之則是低 PD-L1 表達（CT 圖見圖 1.1）；二是 TPS 1%，TPS ≥ 1%為陽性 PD-L1 表達，反之是陰性 PD-L1 表達（CT 圖見圖 1.2）。目前臨床指引也依據 PD-L1 expression 給予治療策略，建議有高 PD-L1 表現（PD-L1 expression ≥ 50%）的患者直接接受 PD-1/PD-L1 抑制劑治療，而較低 PD-L1 表現（PD-L1 expression ≥ 1%）的患者則混和 PD-1/PD-L1 抑制劑與化療藥物進行治療[3]。

雖然 PD-L1 expression 已成為當今廣泛採用的 biomarker，然而 IHC 檢測的精確度卻受到腫瘤檢體和抗體染劑的影響，無論在腫瘤檢體或抗體染劑上，兩者的可靠度目前都尚存爭論。由於 PD-L1 蛋白並非均質地表現，有高度的腫瘤內的不均質性（intra-tumor heterogeneity），但檢體只使用局部的腫瘤取樣而成，檢體的檢測結果難以反應整體腫瘤的 PD-L1 表現，導致容易錯估了 PD-L1 expression[13]。另外在抗體染劑方面，有不同廠牌的抗體、不同廠牌的染色流程、不同的判定標準等等，許多因素介入檢測結果，使的檢測結果缺乏共識[14]。

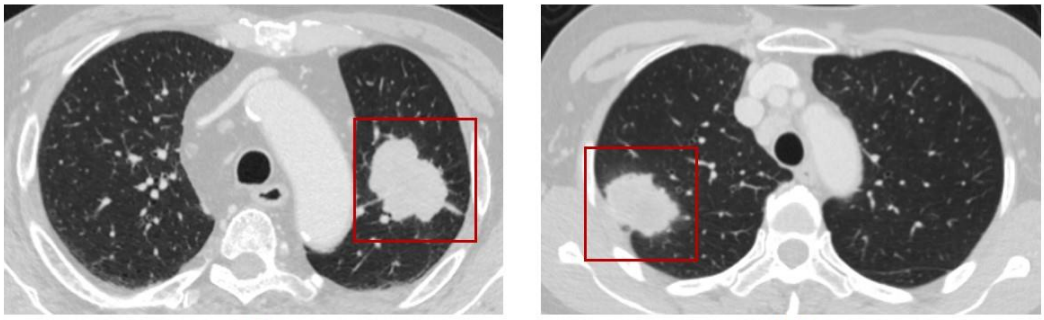

圖 1.1 (a) PD-L1 expression ≥ 50%的 CT 腫瘤影像；(b) PD-L1 expression < 50%的 CT 腫瘤影像

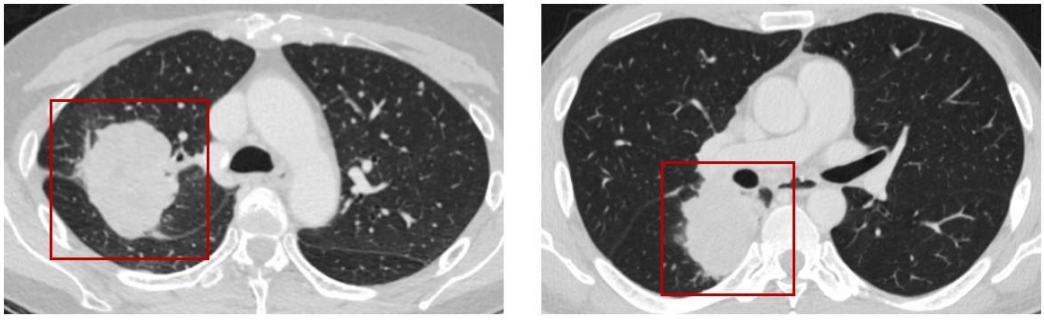

圖 1.2 (a) PD-L1 expression ≥ 1%的 CT 腫瘤影像；(b) PD-L1 expression < 1%的 CT 腫瘤影像

1.3 研究動機與目的

以 PD-1/PD-L1 抑制劑進行的免疫治療為 NSCLC 晚期患者增加了一個治療選項，然而免疫治療的花費昂貴，且仍有過度免疫反應的潛在風險。若給予不適合的患者，除了無法達到治療效果外，也可能導致額外的併發症，與造成不必要的經濟負擔。在治療前篩選出較有可能受益的患者，把握最佳的治療時機，可以讓病患獲得更好的預後狀況。然而，目前判定進行免疫治療與否的 PD-L1 expression 生物標誌檢測方式存在漏洞，由於檢測結果不精確，讓患者的 PD-L1 expression 判定與治療政策無法有統一共識。為了有效幫助免疫治療的患者篩選，需要找到能判別 PD-L1 expression 的 biomarker，並克服現有侵犯式檢查的不足。

為了克服 PD-L1 在檢測上的潛在問題，尋找更準確且穩定的 biomarker 成為重要的關鍵。近年，許多研究希望從影像中找到 PD-L1 expression 的 biomarker，利用不同機器學習和深度學習技術發展之 CAD，從 CT 影像中提取 radiomic features 或 deep features 來預測 PD-L1 表現量。以下將討論目前以影像 biomarker 預測 PD-L1 expression 的發展。

Sun[15]等人的研究是首次使用 radiomics 分析方法來探討多種癌症的免疫表型（immune-phenotype）和影像特徵之間的關聯性，將患者根據腫瘤對免疫的活性，分為 immune-desert、immune-excluded 和 immune-inflamed。結果證實 radiomics 分析所提取的特徵有區分 immune-inflamed 與其他 immune-phenotype 的潛力。Jiang[16]等人提出基於 radiomics features 來評估非小細胞肺癌患者在 PD-L1 的表現量，以 CT 與 PET 影像提取基本 radiomics 特徵，並發現以 CT 影像挑選出的小波（wavelet）特徵，其建立的預測模型表現最佳，在預測 PD-L1 ≥ 50% 的表現為 AUC : 0.91。Bracci[17]等人與 Wen[18]等人透過傳統 radiomics 分析對晚期 NSCLC 患者的 CT 影像做 PD-L1 ≥ 50%的分類，兩組研究團隊都使用 least absolute shrinkage and selection operator （Lasso）來做特徵篩選，並以 Lasso 的係數乘上特徵值線性組合成預測分數，放入邏輯迴歸模型（Logistic Regression）中預測 PD-L1 expression。前者發現 skewness 和 low gray-level zone emphasis （GLZLM）與 PD-L1 ≥ 50%相關，後者發現 kurtosis、cluster tendency（GLCM）、size zone nonuniformity（GLSZM）、gray level nonuniformity normalized（GLRLM），與 wavelet 特徵中的 long run high gray level emphasis 和 high gray level zone emphasis 可以用來預測 PD-L1 表現。Bracci 等人的分類表現 AUC 為 0.789，Wen 等人將上述 radiomics 特徵加上腫瘤及臨床特徵的 AUC 為 0.793。2020 年，Zhu[19]等人提出使用深度學習的模型架構進行 PD-L1 表現量的分類，利用 3D DenseNet（Dense Convolutional Network）來捕捉立體特徵，預測 PD-L1 ≥50%以及≥1%的表現分別為 AUC : 0.765 與 0.78。Tian[20]

和 Wang[21]兩個研究團隊使用相似的架構，結合放射體學、深度學習與臨床特徵來分類 PD-L1 表現量。Tian 等人利用 2D DenseNet 提取深度特徵，radiomics features 則由 Mann Whitney U test 得到 sum entropy （Wavelet-HL-GLCM）以及 gray level nonuniformity normalized （Wavelet-LL-GLRLM），最後使用年齡、性別、抽菸與否和家族癌症史等臨床特徵。綜合上述特徵再經由一層全連接層來分類 PD-L1 ≥50%，得到 AUC 為 0.76。Wang 等人則是將 PD-L1 expression 分成高（≥50%）、中（1%-49%）、低（<1%）三種。使用 3D ResNet（Residual Neural Network）、多種 radiomics features，與治療方法、TMN stage 資訊、病理亞型等臨床特徵，最後綜合 radiomics features、deep features、clinical features 在 PD-L1 expression 高、中、低的表現分別是 AUC: 0.946、0.934、0.95。

綜觀上述研究，從 CT 影像尋找 biomarker，利用電腦輔助診斷對 PD-L1 表現進行分類，有著不錯的效果。然而目前的方法仍面臨到些許挑戰，使用機器學習方法，由於依賴人為特徵來進行分類，容易受到資料的影響，缺乏穩定性與泛用性。相反，深度學習方法可以自行學習特徵，進而用於分類。然而，使用深度學習方法需要大量的訓練樣本，以便模型能夠充分學習。現有的研究中，可用於 PD-L1 表現分類的訓練樣本仍然有限，這可能限制了深度學習方法的表現。

目前對於解決對標記數據量不足的方法中，常見的包括數據增強和遷移學習等等。數據增強通過擴充數據、增加數據多樣性來提高模型的泛化性能，例如變換或是使用生成模型獲取更多樣本。遷移學習通過在大規模數據集獲得的通用特徵提取模型來提高泛化性。近年來，自監督學習能達成使用無標記資料來訓練的架構，對於資料的要求門檻降低。在影像方面，目前主流的遮蓋圖像模型例如 BEiT[22] 以及 MAE[23]，都證實這種利用大量且無標記影像進行自監督預訓練，再遷移至少量的標記資料做微調的方式，相較於傳統監督式預訓練有更好的遷移表現。Masked Image Modeling 利用將部分資訊遮蓋，並訓練模型完成重建，進而使模型

學習提取影像特徵的能力。

本研究的主要目的是解決在使用醫學影像來建立 PD-L1 表現分類 CAD 系統時，因資料量不足而導致深度學習模型過度擬合和無法準確學習特徵的問題（圖 1.3）。為了克服這些限制，本研究旨在探索自監督訓練技術在 PD-L1 分類模型中的應用。為了減輕模型對數據量的依賴性，我們提出了一個基於自監督訓練架構的遮蓋圖像模型，以增強模型的遷移能力。這使得模型可以通過在醫學影像上進行預訓練，從而減少影像特徵之間的差異。同時，為了使遮蓋圖像模型更適應醫學影像特性，避免目標物不明顯的問題，在多任務學習中加入分割任務，使模型在提取特徵時能夠區分前景和背景，能更好地捕捉和學習 PD-L1 表現之特徵。

此外，本研究使用生成對抗網路（Generative Adversarial Networks, GAN）生成的預訓練樣本來增加數據的多樣性。通過利用自監督學習獲得更好的遷移能力，以及生成對抗網路生成的樣本，模型在遷移到下游的 PD-L1 資料集進行訓練時，能提高模型的準確性和穩定性。透過以上方法，本研究希望克服 PD-L1 資料量對分類模型學習的限制，提高影像中 PD-L1 表現的分類表現，進而為臨床提供可靠的在治療策略幫助。

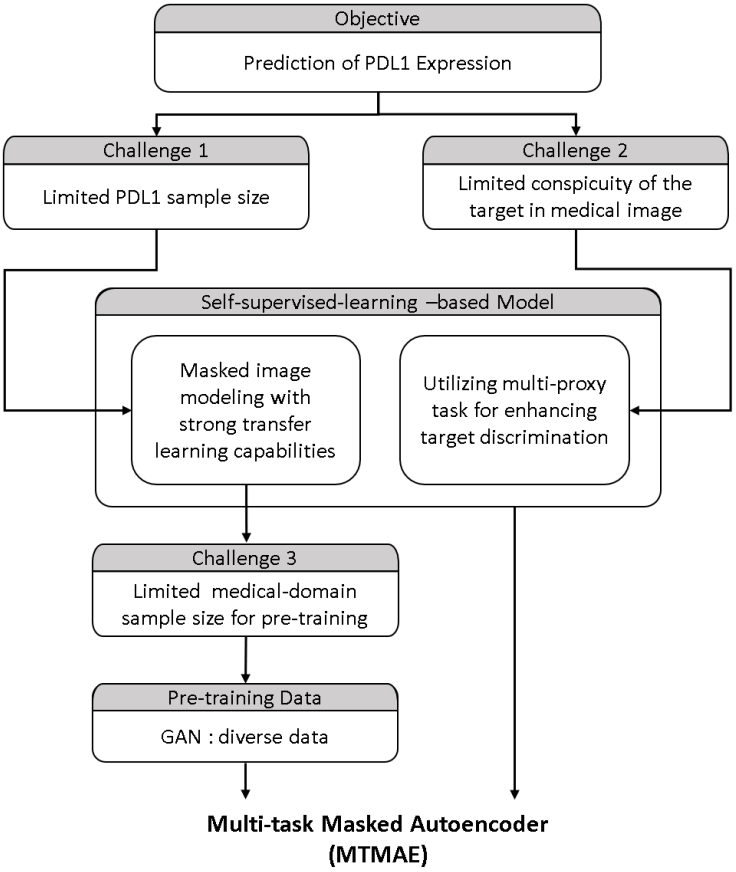
Objective
Prediction of PDL1 Expression
Challenge 1
Limited PDL1 sample size
Challenge 2
Limited conspicuity of the target in medical image
Self-supervised-learning –based Model
Masked image modeling with strong transfer learning capabilities
Utilizing multi-proxy task for enhancing target discrimination
Challenge 3
Limited medical-domain sample size for pre-training
Pre-training Data
GAN : diverse data
Multi-task Masked Autoencoder (MTMAE)

圖 1.3 研究目的

第二章 文獻回顧

本研究的研究目標為使用自監督訓練技術來克服 PD-L1 資料在深度學習時樣本不足，建立能預測 PD-L1 expression 的影像 biomarker，進而幫助患者得到適當的治療計劃。目前，活檢樣本檢測的 PD-L1 expression 變異度高，導致檢測結果不精確。為了幫助患者得到適當的治療選項，找到能基於整體腫瘤判別 PD-L1 expression 的 biomarker 是很重要的。考量到非侵入且能提取完整資訊，影像 biomaker 是許多研究的著手方向，本章節將探討現有使用 CT 影像來預測 PD-L1 表現量的文獻。此外，本研究使用自監督學習的遮蓋圖像模型架構來解決醫學資料不足的問題，也將探討目前遮蓋圖像模型在醫學影像的應用。

2.1 PD-L1 expression 分類

研究證實 PD-L1 expression 與 PD-L1 治療的效果有正相關，且在 PD-L1≥50% 的患者中預後有明顯的改善，臨床對 PD-L1 expression ≥ 50%的患者給予直接接受 PD-1/PD-L1 抑制劑治療的策略。以下將針對利用 CT 影像來預測 PD-L1 表現量≥50%的文獻進行討論，現有文獻主要使用放射體學方法、深度學習，以及臨床資訊來幫助預測 PD-L1 表現量。

2.1.1 放射體學模型

放射體學透過提取影像特徵，結合機器學習的方法形成預測模型，其研究方法可大略劃分為：（1）影像蒐集，通過醫學照影如 CT、PET/CT 和 MRI 取得醫學影像。（2）影像分割，從影像中分割出感興趣的區域，以便針對特定區域提取特徵。（3）特徵提取，針對分割出來的區域做定量影像特徵提取與分析，常見的有形狀特徵、一階特徵，以及紋理特徵 grey level co-occurrence matrices（GLCM）、 grey

level run-length matrices（GLRLM）...等等。（4）特徵挑選，在提取出的特徵中，挑選對任務有鑑別力的特徵，避免多餘的特徵降低模型表現。（5）建立模型，使用挑選出來的有效特徵來建立模型，以此進行分類、回歸等任務。

於 2020 年，Jiang[16]等人首次提出基於 radiomics features 來評估非小細胞肺癌患者在 PD-L1 的表現量，並發現尤其來自 CT 影像的預測模型，可幫助預測患者的 PD-L1 表現。該文獻於 CT 與 PET 影像各提取形狀、一階、二階、wavelet 等特徵共 1744 個，並使用 automatic relevance determination 和 Lasso 挑選各 12 個相關特徵。最後模型以 logistic regression 和隨機森林（Random Forest）來共同預測 PD-L1 的表現量。此文獻評估兩種染色抗體 SP142 與 28-8，並分別對 PD-L1 expression 是否≥50%以及是否≥1%做預測。另外，也比較三種影像輸入的預測表現，分別是 CT、PET，以及 CT/PET，實驗得出以 CT 影像來預測效果最佳。此文獻蒐集的 399 位病人以早期為主，晚期病患約為 25%，並隨機分出 1/3 的資料當成測試級。分類 PD-L1 expression 1%為 AUC: 0.97 與 0.86（SP142/28-8），分類 PD-L1 expression 50%為 AUC: 0.8 與 0.91（SP142/28-8）。

2021 年，Bracci[17]等人用 radiomics features 來對 72 位晚期 NSCLC 病人（stage≥ IIIA）進行 PD-L1 expression ≥50%以及≥ 1%的分類，使用的染色抗體為 SP263。影像由兩位胸腔放射科醫師圈選的腫瘤邊界，並以兩種邊界分別提取了 48 個特徵，包含 16 個形狀特徵和一階特徵，以及 32 個紋理特徵。特徵挑選的方式先挑出兩組邊界特徵中組內相關係數（Intraclass Correlation Coefficients, ICC）≥0.75 的特徵，再使用 Lasso 以及 p-values＜0.10 刪除多餘的特徵。最後，將選中的特徵之 Lasso 係數與特徵值相乘後加總，此值稱為 Rad-score，把 Rad-score 放入 logistic regression 得到預測分數。在分類 PD-L1 expression 是否≥50%的部分，最具代表性的兩個特徵分別為 skewness 和 low gray-level zone emphasis （GLZLM）。隨機抽取 1/3 的資料來生成測試集，分類 PD-L1 ≥50%表現 AUC 為 0.789。分類 PD-L1

<1%的部分則選出 sphericity、skewness、conv_Q3 和 gray-level non-uniformity，AUC 為 0.801。最後將性別、年齡、腫瘤大小等臨床特徵與 Rad-score 一起做統計分析，發現只有 Rad-score 與 PD-L1 expression 有顯著相關（p-values <0.05）。

Wen[18]等人利用 radiomics features 與影像的 morphological factors 來對 PD-L1 expression 與腫瘤基因上的腫瘤突變負荷量（Tumor Mutational Burden, TMB）進行預測。該文獻納入 120 位晚期 adenocarcinoma 患者（stage≥ IIIA），使用的染色抗體為 SP263，並挑出 30 位病患做為測試集。影像由兩位胸腔放射科醫師圈選的腫瘤邊界，並以兩種邊界分別提取形狀、一階、二階、小波（wavelet）等特徵共 462 個。接著，選出 ICC ≥ 0.8 的特徵，再用 Lasso 來挑選有鑑別力的特徵，並建立 logistic regression 模型來預測 PD-L1 expression。在預測 PD-L1 的部分，具有鑑別力的特徵為 kurtosis、cluster tendency（GLCM）、size zone nonuniformity（GLSZM）、gray level nonuniformity normalized（GLRLM），與 wavelet 特徵中的 long run high gray level emphasis 和 high gray level zone emphasis。除了 radiomics feature，在 morphological factors 與臨床特徵統計分析顯示，分化程度以及腫瘤形狀與 PD-L1 表現有關。最後預測 PD-L1 ≥50%於測試集的表現為單獨使用 radiomics features AUC:0.722、單獨使用 morphological factors 與臨床特徵 AUC:0.645，以及兩者合併使用 AUC:0.793。另外，預測 TMB 的表現在綜合利用三種 features 上 AUC 為 0.786。

2.1.2 深度學習與結合其他模型

深度學習的核心特色在於不須先進行特徵提取，由模型在做參數更新時，自己產生合適的特徵。以 ML 的方式做的特徵提取，侷限在人所定義出的定量特徵，此外，模型自我學習的另一優點是可提取更高維度的特徵。以深度學習來進行影像分析的挑戰點在於建構一個能夠達到研究目標的模型，以達到最佳的效果。除了單純使用深度特徵，也有研究結合 radiomics 特徵與臨床特徵一同幫助分類。

Zhu[19]等人使用 3D DenseNet 來提取影像資訊，並應用此模型來預測 PD-L1 表現量。模型通過改良 2D DenseNet 的卷積層和池化層，將其修改為 3D 形式而成，藉此來提取 CT 影像的 3D 特徵。同時先對數據集 ImageNet 進行 transfer learning 來補足資料量較小的問題。該研究收案對象為 127 位晚期 adenocarcinoma 患者，使用的染色抗體為 SP263，將 CT 影像以 VOI 大小 128×128×64 當作輸入，提取 deep features 並進行 PD-L1 表現量≥50%以及≥ 1%的分類。在平均 5 fold cross validation 的結果後，預測 PD-L1 50%以及 1%的表現於測試集分別為 AUC : 0.765 與 0.78。此研究也分析了影像和臨床特徵的關係，發現只有肺轉移與 PD-L1 表現量相關，且與之建立的模型分類能力不如 3D DenseNet。

除了使用 deep features 來對 PD-L1 expression 做預測，Tian[20]和 Wang[21]兩個研究團隊提出了結合放射體學、深度學習與臨床特徵來進行綜合評估。兩者架構相似，皆是對上述三種特徵進行提取後，再將特徵結合，通過全連接層來輸出最後的預測分數。

Tian 等人試圖利用放射體學、深度學習與臨床特徵來預測 PD-L1 expression 與治療預後，試圖多種特徵來互補獲得的資訊。預測 PD-L1 expression 的部分，使用了三種不同的方始提取特徵。首先是深度學習的方式，將影像以 2D slice 的方式放入 DenseNet 來做 deep learning feature 的影像提取。放射體學的部分則是套入許多 filter ，例如 wavelet，local binary patterns(LBP)等等，共提取出 1316 種 radiomics features。經由 Mann Whitney U test 得到兩個具有鑑別力的特徵（Wavelet-HL-GLCM sum entropy，Wavelet-LL-GLRLM-gray level nonuniformity normalized）。最後放入四個臨床特徵，分別是年齡、性別、抽菸與否和家族癌症史。最終結合這三種特徵，放入全連接層來預測 PD-L1 expression 是否≥50%。此文獻納入的資料為 stage IV 的 NSCLC 病人共 939 位，使用的染色抗體為 SP142，並從中隨機分出 1/10 當作測試集，結合三種特徵的方法結果在測試集為 AUC : 0.76。單純使用放射體學方法、

深度學習方法與臨床資訊的 AUC 分別為 0.75、0.68 和 0.48。

Wang 等人將 PD-L1 expression 分成高（≥50%）、中（1%-49%）、低（<1%）三種，並分別對三種表現成度與治療預後進行預測，使用的特徵也是放射體學、深度學習與臨床特徵。預測 PD-L1 的部分，使用 3D ResNet，將影像以 36*36*36 的 VOI 輸入，來提取 deep features。改良傳統的 3D ResNet 把 subsampling 拿掉，來適應較小的輸入。Radiomics features 則提取形狀、一階及二階等紋理特徵共 1672 種，並以 Lasso 與刪除 variance<0.6 的篩選方法，得到 107 個特徵。臨床特徵包含治療方法、TMN stage 資訊、病理亞型等等，並挑選 p-values < 0.05 的放入模型。此文獻的早期病人占了 44%，晚期病人占比約 43%，共 1135 位，使用的染色抗體為 SP142，並隨機分出 1/5 的資料當作測試集，單獨使用 radiomics features 方法來預測三種 PD-L1 表現量在測試集的 AUC 分別是 0.88、0.851 與 0.89（高、中、低）；單獨使用 deep features 的 AUC 分別是 0.901、0.863 與 0.902；最後綜合 radiomics features、deep features、clinical features 的表現是 AUC: 0.946、0.934、0.95。

2.2 遮蓋圖像模型

由於醫學影像資料集通常較小，大多使用自然影像預訓練再遷移應用於醫學影像分析，幫助模型訓練。儘管自然影像與醫學影像存在差異，然而蒐集標記的醫學資料成本高昂，且資料數量有限，要達到自然影像資料集的大小極有難度。自監督預學習的模型，利用大量且無標記進行自監督預訓練，再遷移至少量的標記資料做 finetune。此方法對醫學影像領域而言，大幅降低蒐集資料的難度，並且可以直接使用醫學影像於預訓練中，減少 domain 轉換的差異。以下將針對使用自監督學習架構的遮蓋圖像模型，討論目前在醫學影像分類上的應用。

Zhou[24]等人利用醫學影像來測試 Masked autoencoder (MAE)[23]自監督訓練的有效性，用三種不同的醫學影像任務，包含胸部 X 光疾病分類、腹部 CT 多器官

分割和腦部 MRI 腫瘤分割，來驗證 MAE 的能力。該文獻以 self pre-training 的方式，以相同的資料進行預訓練以及 finetune。實驗結果 MAE 在三種任務上表現皆優於使用 ImageNet 監督式學習預訓練的模型（ViT-B/16），在胸部 X 光疾病分類任務中，AUC 提高了 0.8%；腹部 CT 多器官分割任務中，平均 Dice Similarity Coefficient（DSC）提高了 3.5%；腦部 MRI 腫瘤分割任務中，平均 DSC 提高了 1.13%。另外，與未預訓練的模型（ViT-B/16）相比，三個任務在表現上各提升了 AUC 6.6%、平均 DSC 4.69%和平均 DSC 1.51%。表明自監督訓練在泛化至不同的醫學影像分析任務上有很大的潛力。

Xu[25]使用兩個 CT 資料集（COVID-CT-Dataset、SARS-CoV-2） 進行預訓練方法的分析。比較基於自監督學習方法的 MAE 和其他常見的監督式學習模型（ResNet-101、 DenseNet-121、VGG16、EfficientNetb0 和 EfficientNet-b1），在 transfer learning 能力上的差異。將上述六個模型使用 ImageNet 預訓練後，finetune 在兩個 CT 資料集上。實驗結果表明，使用自監督學習的 MAE 能有效提高模型的泛化能力，表現幾乎與監督學習相同。在兩個資料集中，MAE 在測試集的 AUC 分別為 0.9637、0.9996，僅次最佳模型的 0.9681、0.9997。該文獻也針對 SSL 的預訓練資料集進行探討，發現使用 ImageNet 預訓練的 MAE 相較直接在醫學數據集上進行訓練的 MAE，可以獲得更好的表現。此外將 ImageNet 預訓練好的 MAE 再用醫學影像 pretrain，能減少 domain 差異，表現也有所提升。

最後，Quan[26]等人提出 GCMAE （Global Contrast Masked Autoencoders）的架構，結合 MIM 與對比學習（Contrastive Learning）來學習病理影像的特徵表示。GCMAE 由學習影像重建的 MAE 來負責局部特徵，而 contrastive learning 則是彙集 MAE 學習到的特徵，提取出更有用的全局特徵。研究結果證實在線性分類上 GCMAE 優於 MAE，準確率高出 3.46%。該文獻也將 GCMAE 與監督式學習的 ViT 模型進行比較，GCMAE 準確率上升了 7.32%。由以上實驗可得知 GCMAE 通過結

合 contrastive learning 方法，能夠更有效提取具有高信息密度的病理影像特徵。

綜合以上文獻，分析了 Masked Image Modeling 在醫學影像分析中於分類的應用，也都證明了 Self-Supervised Learning 的有效性。MIM 這種自監督訓練方法在醫學影像分析中的應用具有很大的潛力，可以大大減少樣本需求，從而提高模型訓練的效率和醫學資料的利用。

第三章 模型基礎理論

3.1 Vision Transformer

在自然語言處理（Natural Language Processing, NLP）的領域中，自 2017 年 Transformer[27]這種自注意力機制（self-attention based）的模型被發明後，即成為主流。Transformer 藉由綜觀序列資料中的關係，學習上下文之間的特徵，在 NLP 中有著相當亮眼的表現。由於 Transformer 在 NLP 中優異的表現，2021 年效仿 Transformer 的電腦視覺版本 Vision Transformer（ViT）[28]被提出，ViT 也成功證明在影像辨識上，其成績能超越過去的 CNN 模型。ViT 嘗試使用 Transformer 的結構到影像的領域，並捨棄傳統的 CNN 卷積層結構。CNN 由於卷積層中 sliding window 的作法，只能與 kernel 內的區域做卷積，也就是和相鄰的像素計算相關性。對影像而言，資訊密集度低，相鄰的像素的資訊相近，只對鄰域計算關聯性並不是最有效的做法。由 self-attention 組成的 Vision Transformer 計算每個 patch 之間的關聯性，並從序列中提取重要特徵，把感受野放大到整張影像。

Self-attention module 能學習資料之間的關係，並擷取其中的重要特徵。ViT 模型主要由三個區塊組成 Embedding、Transformer encoder 與分類層所構成，如下圖 3.1，以下將對各區塊做介紹。

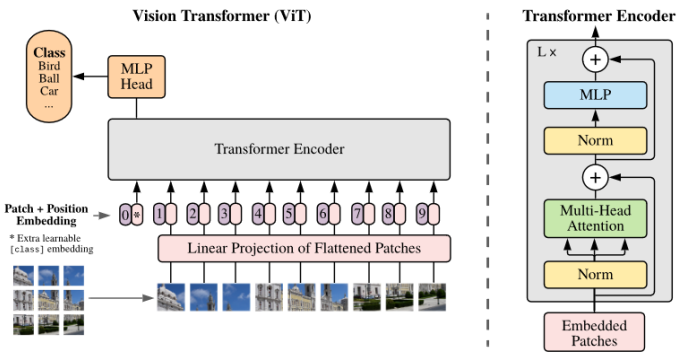

圖 3.1 ViT 架構圖[28]

3.1.1Embedding

Embedding層的主要目的是將圖像預備成Transformer架構的輸入。Transformer是一個 Sequence to Sequence 的網路，輸入序列可看作二维的矩陣（N, D），其中 N 是 sequence 的數量，D 是 sequence 中每個向量的维度，而每個向量又可稱為 token。相對於 NLP 領域中輸入單位為一個詞彙，ViT 則提出將圖像分割成圖塊（patch），以 patch 為單位，如同一個詞彙，接著再轉換成序列的方式作為輸入。每個輸入 token（z^0）由三個部分組成：patch embeddings、class embedding 與 position embeddings。

首先，將圖像 x 拆分成 N 個大小為（P^2C）的小塊 patch，其中（H，W）是原始圖像的大小，C 是 channel 數，（P，P）是每個 patch 的大小，N＝HW/P^2是總 patch 數量。接著，將每個 patch 線性投影至一維向量（D）中，D 的維度是P^2C。以上將影像轉為 patch 再投影至一維向量的輸出稱為 patch embeddings，如式(1)。以常見的 ViT 輸入大小為例，一張 224*224*3 的影像，與一個 patch 原始圖像大小 16*16*3。圖像將被拆分為（224/16）^2=196 個 patch，每個 patch 會被投影成維度是 16∗16∗3=768 的向量，把 196 個 token 疊加在一起後维度就是[196, 768]。

$$\mathrm{x} \in R^{\mathrm{HWC}} \rightarrow \mathrm{x}_p \in R^{\mathrm{N}(P^2\mathrm{C})} \qquad (1)$$

Class embedding 被放在 embeddings 的最前面，任務是用來學習類別的資訊，class embedding 會與序列中所有的 token 做運算，獲得全部 patch 的相關資訊。Position embeddings 是一個加在 patch embeddings 上的一維向量，用來保留位置資訊，以此獲得區分區域的能力。

3.1.2Transformer encoder and Classification layer

Transformer Encoder 是主要提取特徵的架構，透過堆疊數個 Transformer

Encoder block 來實現，每個 encoder layer 包含：multi-headed self-attention（MSA）、multi-layer perceptron（MLP） 和 layer normalization（LN）。依據架構中參數量的不同（encoder block 數量、embedding 大小、MLP 神經節點數，以及 multi-attention 數量），ViT 模型可分為大小不同的變形，見下表 3.1。

表 3.1 ViT 模型參數

Model	Encoder layers	Hidden size (D)	MLP size	Attention heads
ViT-Base	12	768	3076	12
ViT-Large	24	1024	4096	16
ViT-Huge	32	1280	5120	16

3.1.2.1 Multi-headed self-attention

Self-attention 對輸入序列中的每個 token 做 attention 計算，使得模型能夠在提取特徵時找到相關性高的 token 進行重點考慮，而不是對每個輸入 token 做同樣的處理。Self-attention 依照序列中 token 之間的相關性計算 attention 權重（attention weights） ，對其進行加權輸出。權重越大則表示越聚焦於該 token 取出的資訊，藉此在大量輸入中找出重要訊息，而每個 token 對應的輸出都是考慮過整個輸入序列的關聯性得到。

圖 3.2 為 Self-attention 的運作方式，首先對輸入（X）進行矩陣運算，得出 query（Q）、key（K）、value（V）三個矩陣。使用內積來計算 query 和 key 的相關性，再除以一個常量值（dk）並通過 Soft-max，讓值呈現機率分布，計算出的值即 attention weights。最後將 value 和對應的 attention weights 相乘得到輸出（Y）。此運算可以表示為式(2)。

Multi-headed self-attention 是結合多個 self-attention 結構。每個 head 擁有自己

的 query、key、value 矩陣，使其能學習到不同的相關性、不同的特徵，為模型帶來更大的學習空間。最後將每個 head 的輸出相接經過矩陣轉換(W）得到輸出，可以表示為式(3)，其中 n 為 head 的數量。

$$\text{Attention}(Q,K,V)=\text{softmax}(\frac{QK^T}{\sqrt{d_k}})V \quad (2)$$

$$\text{MultiHead}(Q,K,V)=\text{Concat}(\text{Attention}_1, ..., \text{Attention}_n)*W \quad (3)$$

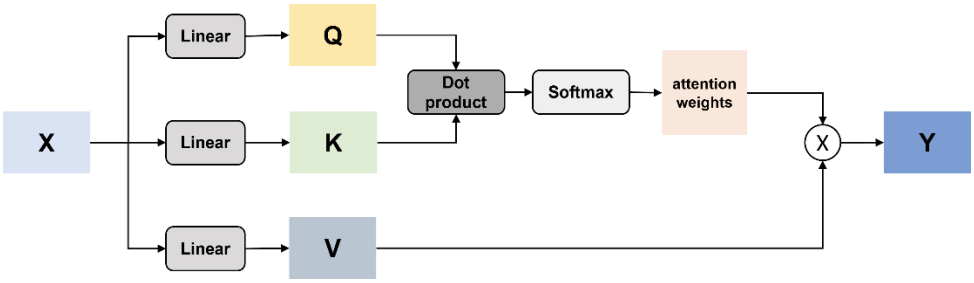

圖 3.2 Self-attention

3.1.2.2 Layer Normalization

Normalization 的目的是讓特徵更穩定，讓特徵以常態分佈落在一定的範圍內，也有正規化(regularization)的功能。CNN 模型常使用 Batch Normalization，但 Batch Normalization 是照一個批次的樣本數（batch size）來計算，當 batch size 較小時，就難以表達全部樣本的分布。Layer Normalization 則是對單一個樣本做標準化，解決了 Batch Normalization 在推估樣本分布的難題，也讓樣本間沒有依賴關係。Layer Normalization 對每個樣本（x）的特徵計算平均值（μ）和標準差（σ），並對其標準化如式(6)，其中 H 是隱藏層節點的數量。此運算可以表示為式(7)，γ和β是一組可訓練參數，為了使特徵經過 LN 不被破壞。

$$\mu = \frac{1}{H}\sum_{h=1}^{H} x_h \quad (4)$$

$$\sigma = \frac{1}{H}\sum_{h=1}^{H} (x_h - \mu) \quad (5)$$

$$\hat{x} = \frac{x_h - \mu}{\sigma} \quad (6)$$

$$LN(x) = \gamma\hat{x} + \beta \quad (7)$$

3.1.2.3 Multi-layer perceptron and GELU

多層感知器（Multi-layer Perceptron，MLP）是一種由多層節點組成的神經網絡，將輸入向量映射到輸出向量。MLP 以全連接的方式層層相接，使用反向傳播算法進行學習，以減少訓練過程中的偏差。MLP 由於節點帶有非線性的激活函數（activation function），能改善網路線性的輸出的問題。常見的激活函數包括 sigmoid,、ReLU、Leaky ReLU 等，ViT 模型選用的 activation function 為 Gaussian Error Linear Unit （GELU），圖形如圖 3.3。GELU 融入 Dropout 與 ReLU 的想法，希望把不重要的資訊歸零。GELU 如式(8)對於輸入 x 乘上一個伯努利分布的 m，即 m~ Bernoulli(Φ(x))，其中 Φ(x) = P(X ≤ x)又 X~N(0,1)，是一個常態分布的累積分布函数，可以誤差函數（erf）表達。此設計可按照當前輸入比其它輸入大多少來進行縮放，隨著 x 的降低，Φ(x)也跟著降低，被歸零的機率就會變大。GELU 以標準高斯分布 N（0, 1）可用 tanh 得到近似函數如式(9)，不少模型使用此近似來實現，也可以將高斯分布的均值和變異數當作參數訓練。

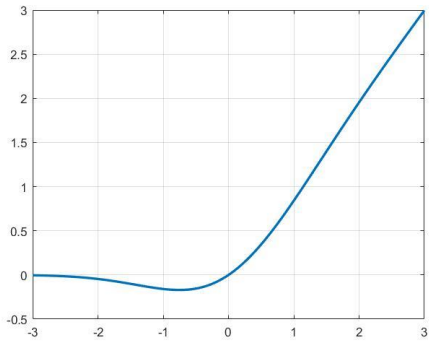

圖 3.3 GELU

$$\mathrm{GELU}(x) = xP(X \le x) = x\Phi(x) = 0.5x(1 + \mathrm{erf}(\frac{x}{\sqrt{2}})) \quad (8)$$

$$\mathrm{GELU}(x) = 0.5x\left(1 + \tanh\left[\sqrt{\frac{2}{\pi}}(x + 0.044715x^3)\right]\right) \quad (9)$$

3.1.2.4 Classification layer

最初的 Vision Transformer 設計為了貼近 BERT 的作法，若下游任務為分類，則會取出 Transformer Encoder 產生的 class token。這個 class token 是在 embedding 階段為了學習類別所放入的 token，可看作是對類別提取的特徵，再接上 MLP 或是一層全連接層來輸出最終結果。另外，關於最後的提取特徵，另一種常見的做法是將 Transformer Encoder 的輸出特徵做全局平均池化（global average pooling），得到全局特徵後做分類。最後一層的全連接層的神經元個數即為分類類別數量。

第四章 研究方法

4.1 研究材料

本研究所使用的資料來源為台大醫院、台大醫院新竹分院，以及台大醫院雲林分院。納入對象為經過臨床病理檢查為肺癌（經手術後檢體或微侵襲性切片證實），並且肺部腫瘤檢體經 IHC 檢測過有 PD-L1 expression 資料之病患。排除條件為檢測 PD-L1 expression 前，已有治療紀錄（如放射治療、化學療程、手術或免疫治療等），以及腫瘤複雜邊界難以定義之患者。本研究總共蒐集 188 例病患，其中 PD-L1 expression ≥ 50%(+)有 49 例，PD-L1 expression < 50% (−)有 139 例。所收案之病患臨床資訊包括年齡、性別、病理分類以及肺癌分期，列於表 4.1。臨床資訊中於 PD-L1 expression 的分類並無顯著差異，除了性別上較有相關(p-value=0.065)。

由於本研究有多個 IHC 檢測抗體，需統一檢測標準才可獲得一致的 PD-L1 表現，根據[29]在針對 NSCLC 上對不同試劑上的研究結果，研究表示 22C3 與 SP263 表現一致，SP142 則傾向有較低的表現。因此，本研究在試劑上優先採用 22C3 的結果、若無則使用 SP263，或 SP142 檢測表現為 0%的患者。在收案的 188 位患者中，使用 22C3 的共 112 位、使用 SP263 的共 74 位，以及 2 位為 SP142 PD-L1 表現為 0%。

本研究的 CT 影像使用患者 IHC 檢測前的 contrast-enhanced CT， 影像細切 slice thickness 介於 0.625 ~ 1.25mm。CT 設備為 GE MEDICAL SYSTEMS、SIEMENS 以及 Philips。採用參數如下：單一切面大小為 512 x 512 pixel；Tube voltage : 100 ~ 140 KVp；X-ray tube current : 57 ~ 849 μA。

表 4.1 收案病患臨床資訊

	Total	PD-L1 ≥ 50%	PD-L1 < 50%	p-value
Age	67	66	68	0.238
Gender				0.065
M	105	36	69	
F	83	13	70	
Histopathology				0.781
ADC	126	28	98	
SCC	30	8	22	
Others	32	13	19	
Stages				0.246
I	23	3	20	
II	11	2	9	
III	30	11	19	
IV	124	33	91	

4.2 研究方法

4.2.1影像處理

在過去直接使用 GAN 來生成訓練資料，直接用於分類、診斷等，往往有資料能否代表真實影像數據的爭議，尤其在醫學影像領域。因此，本實驗僅使用 GAN 生成之影像作為預訓練資料，提取高品質與高多樣性的影像特徵。GAN 生成之預訓練資料為大小 64*64*64 的 VOI，以及作為生成條件的結節標記區域（mask）同樣大小為 64*64*64，像素間距為 1mm。考慮到結節在影像上的占比，僅使用結節體積大於 3300mm^3 之樣本。為了充分使用肺部資訊，使用 lung window (Window width:1500、Window level:524)，再進行正規化。將 3D CT 資料轉成 2D 影像當作輸入，在 VOI 上使用 axial plane 挑選腫瘤出現的 slices 並且間隔取樣，並將取出的 slices 放大輸出成 224*224 的 2D 影像。最後，以 224*224 的 2D CT 影像以及其 2D

標記區域作為輸入，流程如圖 4。

至於下游任務的 PD-L1 資料，首先將像素間距線性內插調整為 1*1*1mm，接著依照所收集腫瘤樣本的大小，以腫瘤為中心設置大小為 128*128*128 的 VOI。PD-L1 資料同樣使用 CT 影像的 lung window，再進行正規化。由於分類模型使用 2D 影像當作輸入，因此會使用 axial plane 挑選腫瘤出現的 slices。在驗證集與測試集中會取出每個 case 腫瘤面積最大的 slice，再將該張影像切下 120*120 的區域 resize 至 224*224。訓練集部分，則與預訓練資料相似，會在腫瘤出現的 slices 中做間隔取樣，並同樣將影像切下 120*120 的區域 resize 至 224*224。訓練集中為了平均兩個類別的影像數量，兩個類別的間隔大小與兩個類別的樣本比例相同。如此兩個類別最終取出的 2D 影像數量會相當，流程如圖 4。

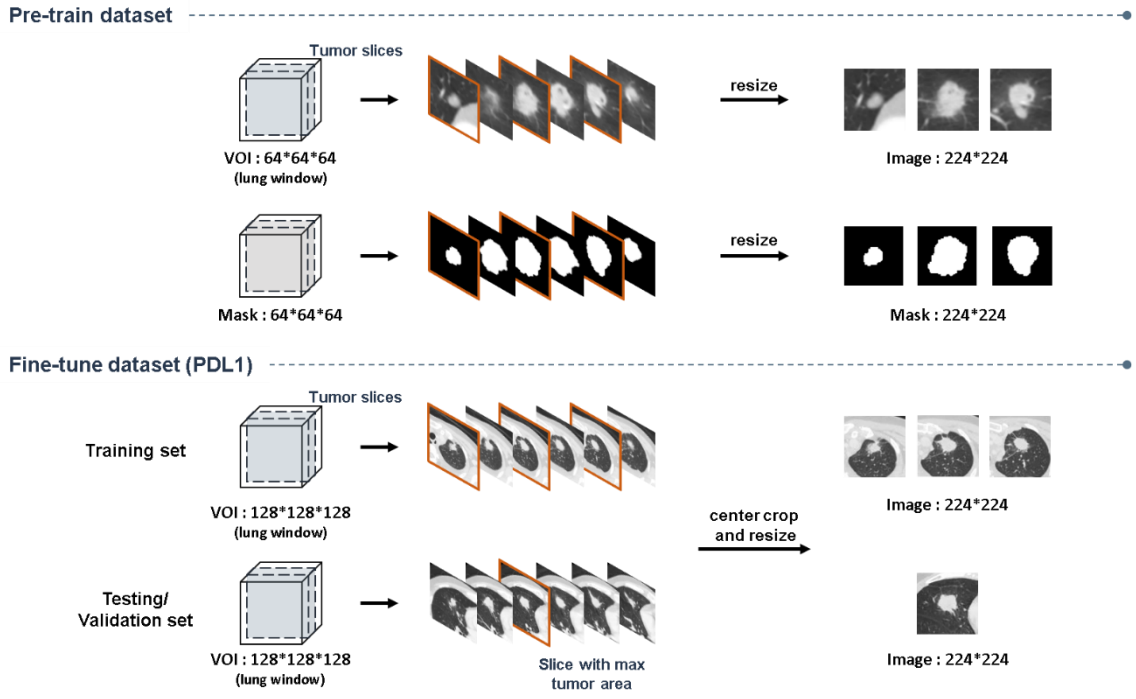

圖 4 影像前處理流程圖

4.2.2遮蓋圖像模型

隨著硬體條件的提升，更多龐大的模型被提出，這些複雜的模型能夠處理更加複雜的任務並展現更好的表現。然而，這也意味著需要更多的訓練數據，標記數據不易取得，這使得數據量成為目前模型性能的一個重要限制因素。近年來，自監督學習在自然語言處理領域成功地解決了大型 Transformer 模型面臨的數據限制問題，例如使用 Masked Language Model（MLM）架構的 BERT。這種基於遮蔽技術的自監督學習預訓練模型的核心概念是將部分數據遮蔽並學習預測被遮蔽的內容，然後再使用較少的標記數據進行特定任務的微調（fine-tune）。

過去，在計算機視覺領域中，主流模型一直是能夠突出空間特徵的卷積神經網絡，並搭配監督式預訓練的方法。然而，隨著 Vision Transformer 的提出，該模型證明了其在性能上能超越傳統的 CNN 模型。為了跟進自然語言領域中的自監督學習預訓練方法，許多研究團隊開始嘗試複製 MLM 的概念，並將其應用於計算機視覺領域，從而設計出適用於影像的 Masked Image Model（MIM）。例如，Bao[22]等人提出了 BEiT，將 BERT 的 SSL 預訓練方法複製於影像領域，成功地開發出一種適用於影像的 MIM 模型。同樣地，He[23]等人提出了 Masked Autoencoder（MAE），使用 autoencoder 在同一個模型中學習圖像特徵與還原，提高了模型的效率。此外，還有更為精簡的 SimMIM 模型[30]，其中只包含一個 MLP 解碼器。上述的 MIM 模型已經證明，相比於監督式預訓練，使用自監督學習的方法在遷移學習方面有更好的能力並且有更廣泛的應用性，能夠有效提升下游任務的表現。

遮蓋圖像模型的訓練過程包括預訓練和微調兩個階段。在預訓練階段，模型對圖像進行隨機的遮蓋，並要求模型預測被遮蓋的區域，通過比較預測結果和實際圖像來計算損失函數。透過遮蓋預測的任務，模型能夠學習圖像中的特徵和紋理，從而提高對圖像的理解和表示能力。預訓練階段完成後，進行微調階段，使用少量有

標籤的圖像資料和特定下游任務（如圖像分類、分割或物體偵測）的標籤資料進行模型微調。這將預訓練模型的參數作為初始權重，將預訓練的知識轉移到下游任務上，以進一步訓練模型。微調的目的是將預訓練階段學習到的特徵和資訊應用於具體任務，提高模型在下游任務中的表現。這種訓練方法使得模型能夠利用大量的未標記資料充分地理解和表示圖像，同時也能夠在資料有限的情況下適應特定任務。

本研究使用遮蓋圖像模型中的 Masked Autoencoder 架構，並針對醫學影像的特性對其加以改良，以下將此模型進行介紹與其架構的說明。

4.2.2.1 Masked Autoencoder

Masked Autoencoder 使用去噪自編碼器（Denoised Autoencoder）為主要架構，autoencoder 是一種學習表徵（representation）的非監督式學習方法，由兩個神經網路組合而成，一個作為編碼器（encoder），另一個則是解碼器（decoder）。Encoder 將輸入投射到潛在表徵，decoder 再負責把輸入重構，目標是讓輸出和輸入越接近越好。Denoised Autoencoder 利用加入噪音或是遮蓋的方法，先破壞輸入再放入 encoder，強迫學習原始輸入的潛在表徵，以及重建的抗噪能力。

由於影像資訊密度與文字的差異，影像相較於文字資訊密度較低，單一像素並無太大的語意特徵，一定面積的圖塊才能表達資訊。因此，MAE 的重建任務需要克服 MIM 中的兩大問題，一是讓模型學習到 semantic 特徵，二是讓重建任務無法單靠插值完成。考慮到上述問題，MAE 將輸入圖像切割成圖塊（patch），以 patch 為單位讓 encoder 學習，讓學習到的潛在特徵有意義。接著在遮蓋時，選擇非常高的遮蓋率（75%），確保這個重建的非監督式學習任務是有難度的，無法單靠相鄰 pixel 來填值。

MAE 流程如圖 4.5 所示，在預訓練階段， MAE 將圖像分割成大小相同的

patches，並以 patch 為單位做隨機遮蓋。接著將未被遮蓋的 patches 放 encoder 當成輸入，讓 encoder 能專注學習影像特徵。最後，decoder 接收 encoder 輸出的 unmasked patch features 和以 masked token 形式呈現的 masked patch，來執行重建任務。Decoder 藉由預測 masked patch 的像素值來重建影像，輸出的向量即為每個 patch 的像素值。在 fine-tune 階段只使用 encoder，並將完整影像（不進行 masking）一樣切成 patch 以序列的方式作為輸入。經由 encoder 學習完的輸出特徵，再依照下游任務的性質做利用。

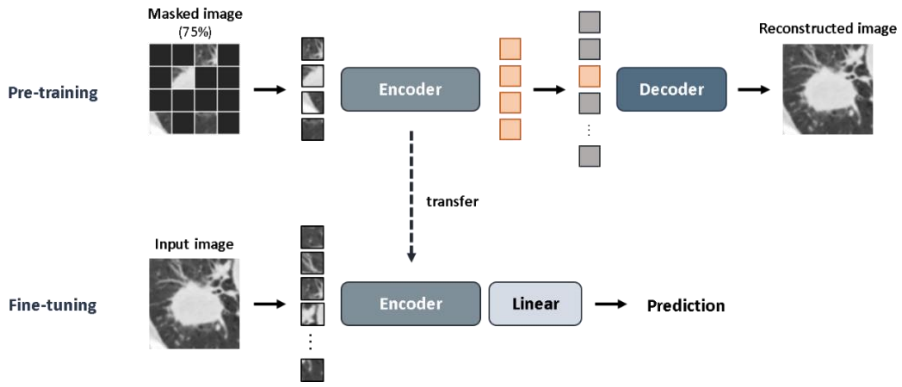

圖 4.5 Masked Autoencoder（下游任務以分類為例）

4.2.3 Multi-task Masked autoencoder (MTMAE)

本研究針對醫學影像的特性對 Masked Autoencoder 進行了改進，提出了一名為 Multi-task Masked Autoencoder（MTMAE）的架構。與自然影像不同，肺部 CT 影像中的目標腫瘤大小各異，且與周圍的血管、器官等組織具有相似的灰度值，因此難以進行準確的區分。然而，如果僅僅去除背景，將會喪失有用的資訊。因此，我們的目標是讓模型更專注於腫瘤區域，同時能夠區分腫瘤與背景。本研究提出的 MTMAE 採用多任務學習的方法，同時訓練模型進行影像重建和影像分割兩個任務。通過引入分割任務，模型可以獲得更多關於腫瘤和背景的資訊，從而實現更精確的腫瘤定位和分析。

MTMAE 的架構如圖 4.6 MTMAE 架構圖所示，包含一個 Encoder 和兩個 Decoder。輸入影像包含腫瘤 CT 影像以及其二元標記區域，其中只有 CT 影像會進入模型中。兩個 Decoder 分別負責影像重建和影像分割任務，並通過 Encoder 與其連接。最後重建以及分割出的影像，會分別與輸入影像計算損失。透過共享 Encoder 的設計，Encoder 在學習影像的語意特徵時也能夠將腫瘤區域和非腫瘤區域的資訊納入考量。相較於僅學習重建任務，這樣的分工使得 Encoder 在轉換至下游任務時對圖像有更好的理解能力。在 MTMAE 中，Encoder 和兩個 Decoder 都採用了 ViT-Base 作為主要架構，由多個 ViT encoder block 堆疊而成。

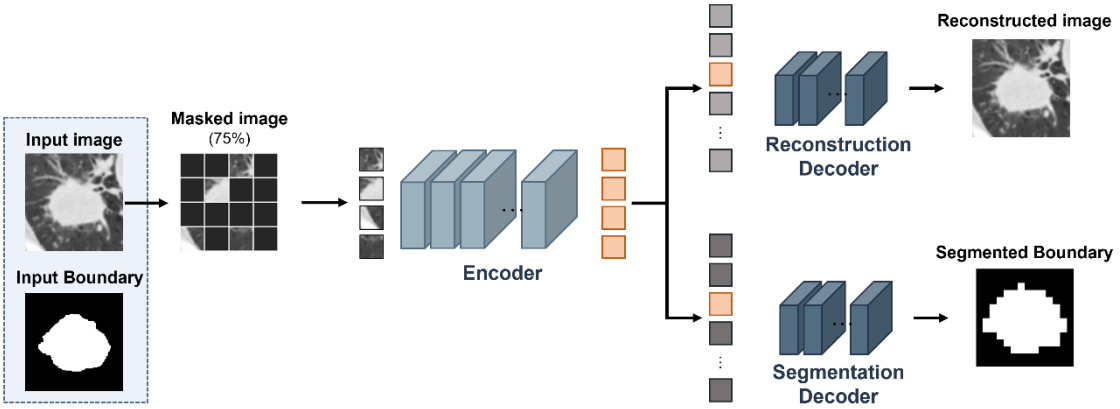

圖 4.6 MTMAE 架構圖

Encoder 的主要任務是學習影像的特徵。它的架構包括 encoder embedding layer 和 12 層的 ViT encoder block，如圖 4.8 所示。Encoder embedding layer 將輸入影像切割成大小為 16x16 的圖塊，並將其轉換成序列形式進行嵌入。每個序列向量的維度為 768（embedding dimension）。接著，對 75%的圖塊進行遮蓋，只保留未被遮蓋的圖塊作為 Encoder 的學習對象，也就是下一個 encoder block 的輸入。

每個 encoder block 的輸入和輸出都保持相同的維度。Encoder block 首先經過

Layer Normalization，然後是 Multi-headed self-attention layer（MSA）。MSA 使用 12 個 multi head，幫助提取不同的特徵。MSA 的輸出與 Eencoder embedding layer 相加，達到殘差連接（Residual Connection）的效果。接下來再進行一次 Layer Normalization，然後通過 Multi-layer perceptron layer（MLP）。MLP 由兩個全連接層組成，中間使用 GELU 激活函數，幫助模型學習非線性特徵，兩個全連接層的大小分別為 3072 和 768 個節點。最後的輸出是將 MLP 的輸出與第二次正規化之前的輸入相加，即第二次的 residual connection。當模型的深度增加時，梯度很容易消失或爆炸，這會使得模型難以訓練。Residual connection 可以幫助梯度更好地流通，使得模型更容易訓練。每個 ViT encoder block 的流程可以表示為式(10）和式(11)，詳細流程見圖 4.4。

$$x'_l = \mathrm{MSA}(\mathrm{LN}(x'_{l-1})) + x'_{l-1} \quad (10)$$

$$x_l = \mathrm{MLP}(\mathrm{LN}(x'_l)) + x'_l \quad (11)$$

$$\text{Dice Loss} = 1 - \frac{2|A \cap B|}{|A| + |B|} \quad (12)$$

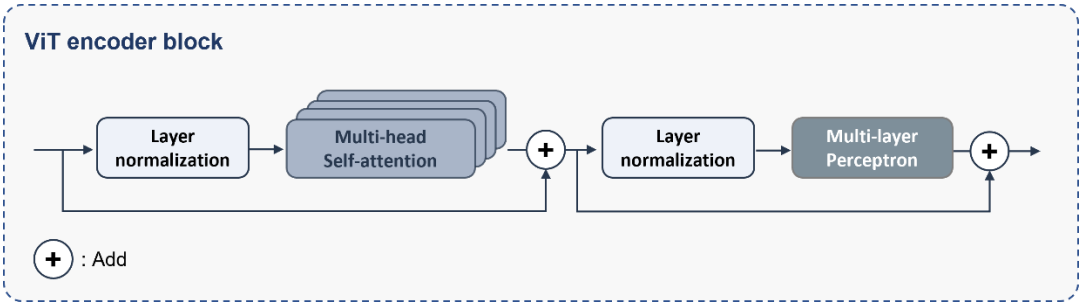

圖 4.4 ViT encoder block

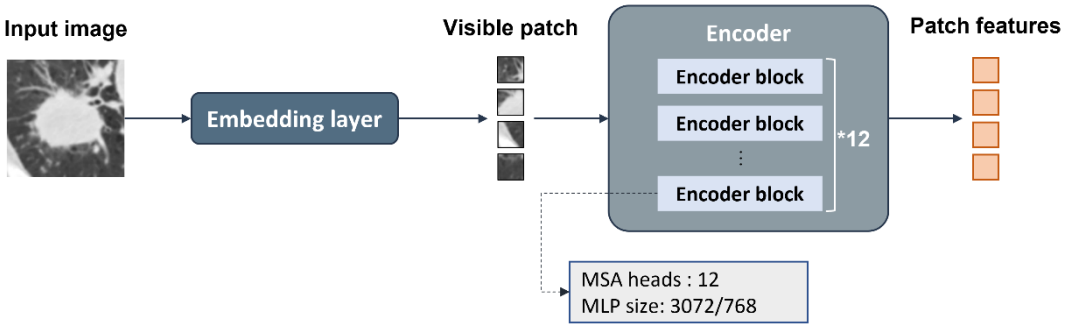

圖 4.8 Encoder 架構圖

兩個 Decoder（Reconstruction Decoder 以及 Segmentation Decoder）的架構相同，包括 decoder embedding layer 和 8 層的 ViT encoder block，如圖 4.6 所示。Decoder embedding layer 負責處理兩種輸入。首先，將 Encoder 輸出的圖塊特徵經過一個全連接層投影到 512 維，也就是 Decoder 的 embedding dimension。其次，給 masked patches 一個可學習的 masked token，作為初始化的 embedding features。最後，將這兩者按位置排列，作為後續 decoder block 的輸入。

Decoder block 的架構與 ViT encoder block 相似，不同之處在於 Decoder block 的 MSA head 數量為 16，以及 MLP 中兩個全連接層的大小分別為 2048 和 512 個節點。最後，Reconstruction Decoder block 的輸出會再經過一個全連接層投影到 768 維，即圖塊的長度乘以寬度乘以三個通道的乘積，投影後的值即為重建解碼器重建的像素值。另一方面，Segmentation Decoder block 的輸出則會經過一個全連接層投影到 1 維，以圖塊為單位預測該圖塊是否為腫瘤區域，結果即為分割解碼器分割出的二元分割遮罩。

損失函數將兩個任務各自的損失（loss）相加。重建任務使用均方誤差損失（mean square error loss）在像素級別計算被遮蓋的圖塊原圖與 Reconstruction Decoder 重建後圖塊的平均平方誤差，以訓練重建出的影像與原圖越相近越好。在

分割任務中，首先將腫瘤標記區域以圖塊為單位取最大值，即若該圖塊包含任何腫瘤區域，則將其視為腫瘤區域。如此可以充分利用腫瘤邊界的特徵。使用 Dice Loss 作為損失函數，計算 Segmentation Decoder 預測的遮罩與腫瘤分割遮罩之間的差異，如式(12) 所示，其中 A 代表分割結果的區域，B 代表真實腫瘤分割的區域。

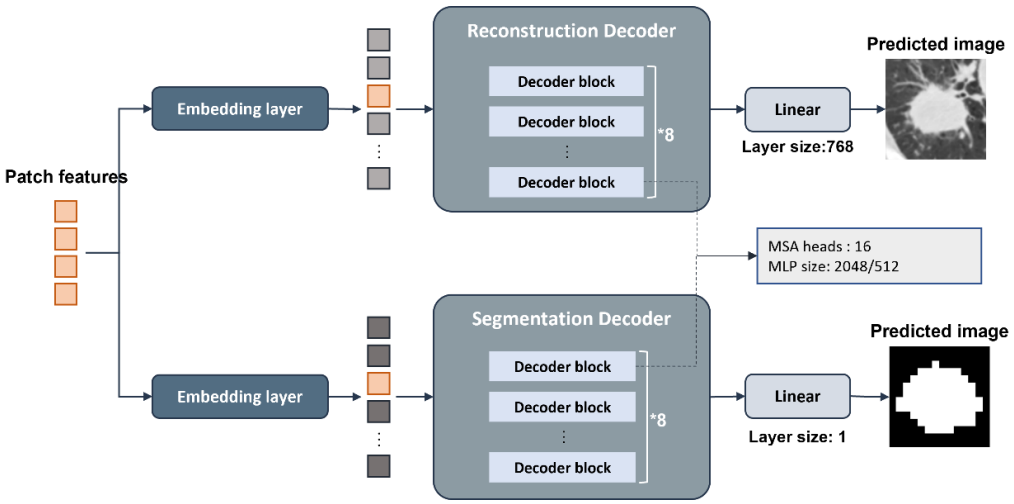

圖 4.6 Decoder 架構圖

在 Fine-tune 階段，我們僅使用 MTMAE 模型中的 Encoder 並將其權重進行轉移（transfer）。Fine-tune 架構如圖 4.7 所示，將 Encoder 的輸出特徵進行全局平均池化，以獲得全局特徵。然後，添加一個全連接層作為分類器，並使用 sigmoid 函數將輸出分為 PD-L1（+）和 PD-L1（-）兩個類別。

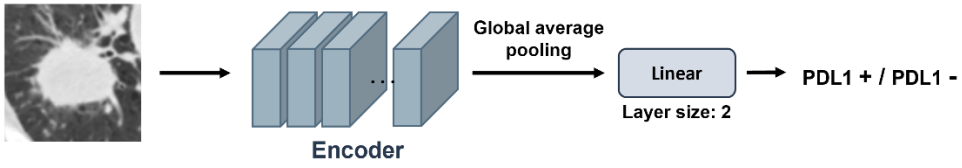

圖 4.7 Fine-tune 架構圖

4.2.4 生成對抗網路模型生成影像

深度學習需要大量的數據來學習有效的特徵。儘管 Masked Image Modeling 可使用無標記影像進行 Self-Supervised Learning，然而即使是無標記醫學影像，數量仍不及 ImageNet 這樣的大型數據集可以提供足夠的數據來訓練深度神經網絡。使用傳統的數據增強（Data Augmentation）如位移、翻轉等等，與原資料仍高度相關，難以產生足夠的多樣性。

為了解決這個問題，本研究透過實驗室先前所開發之生成對抗網路模型[31]（Generative Adversarial Nets, GAN），來生成擬真的醫學影像，並作為預訓練的資料來源。通過 GAN 可以達到增加樣本，並且同時具有特徵多樣性，幫助解決醫學影像不足的問題，提供更廣泛、有用的數據。GAN 由兩個互相對抗的神經網絡組成：生成器（Generator）和判別器（Discriminator）。生成器的任務是從隨機編碼中生成與真實數據相似的數據，而判別器則是訓練為區分真實數據和生成數據。這兩個網絡不斷相互學習，以使 Generator 生成的數據更接近真實數據，同時讓 Discriminator 更能夠區分真實和假的數據，通過這種方式，讓 GAN 可以產生擬真的數據。

實驗室先前所開發之 Gabor-loss GAN 模型，使用公開的結節資料庫 LIDC-IDRI（Lung Image Database Consortium and Image Database Resource Initiative）生成擬真肺結節樣本。為了控制肺結節生成大小及形狀，利用結節分割的標記來引導生成。大致流程為肺結節分割及背景提取、訓練 GAN 模型，以及生成肺結節樣本，流程圖如圖 4.8。首先，會分割肺結節與肺實質背景，接著將背景與二元的分割 mask 作為輸入，訓練 GAN 生成肺結節樣本。Gabor-loss GAN 模型在 Generator 與 Discriminator 皆使用 CNN 架構。訓練 GAN 的 Generator 過程中，除了 Discriminator 給予真實或生成樣本的真偽分類損失，為了增強結節內紋理特徵的生成，會將生成

樣本透過 Gabor filter 進行濾波，並對濾波特徵的整體特徵強度差異進行損失計算（Gabor-loss）。加入 Gabor-loss 可以避免模型過度擬合而導致與訓練樣本過於相似。

訓練完 GAN 模型後，需要為其提供適當的肺實質背景作為輸入條件來生成樣本。為了產生適當的肺實質背景，使用本實驗室先期開發之肺部組織分割演算法[32]，對肺區進行分割，然後進一步分割肺部內的氣管與血管，以及原始資料中的結節區域。獲得適合用於生成結節資料的肺實質區域後，隨機選取一個大小為 64*64*64 的 VOI 作為肺實質背景，並將真實結節標記區域放置在該 VOI 中間。為了確保背景合理性，如果標記區域重疊到肺壁、心臟、氣管等區域，將捨棄該部分重疊的標記區域。最後，若剩餘的標記區域體積小於原有的 80%，則不採用此肺實質背景。完成前處理即可將肺實質背景 VOI 以及結節標記區域給 GAN 生成樣本。為避免生成結節與背景不自然的邊界，使用高斯濾波器平滑處理以及 region filling 來處理連接處，得到最終的生成樣本如圖 4.9 所示。

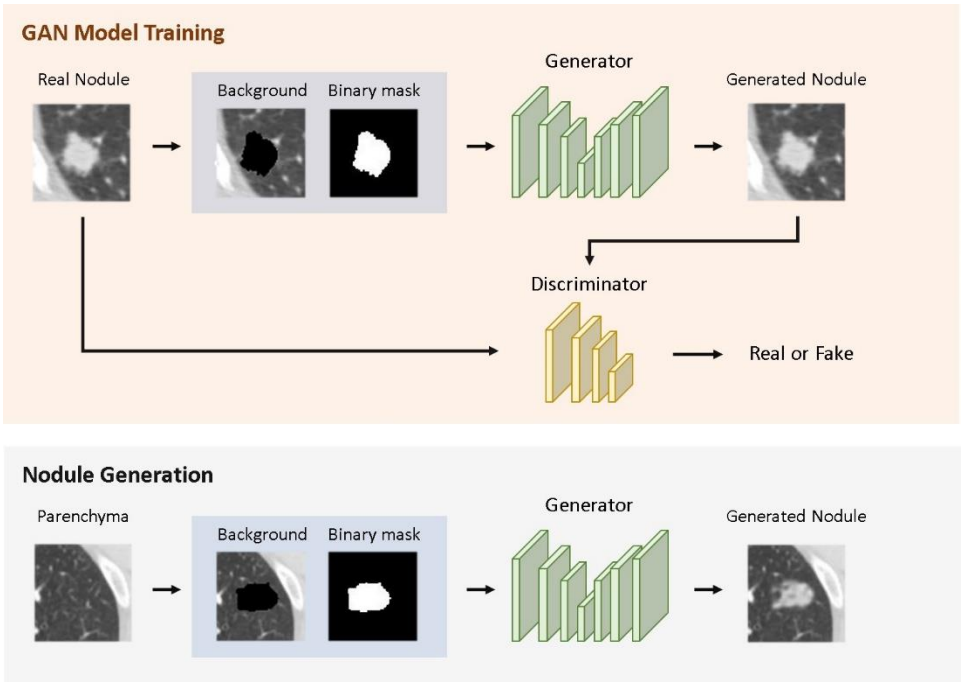

圖 4.8 Gabor-loss GAN 模型流程圖

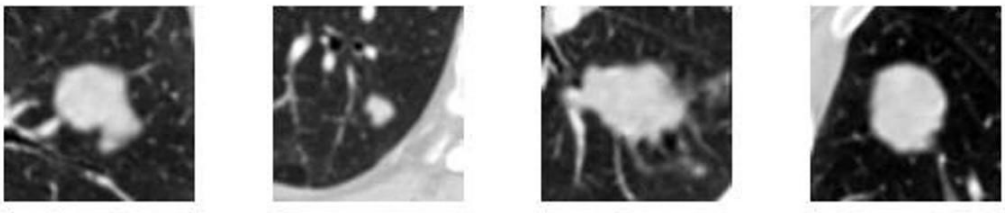

圖 4.9 由 Gabor-loss GAN 生成的肺結節樣本

4.2.5Attention Visualization

相較於 CNN 使用梯度變化或是最後一層特徵圖來獲得分類決策的資訊。Transformer 模型較常使用注意力權重（attention weights）來解釋模型關注的範圍。MTMAE 使用 self-attention 機制來計算各區域的關聯性，並以此獲得重要特徵。本節嘗試探討 MTMAE 學習到的特徵，利用[33]所提出之 Attention Rollout 方法來對模型關注的區域進行可視化。

Self-attention layer 中輸入序列的特徵在 patch 之間進行結合，接著往下傳遞。Attention Rollout 依照 self-attention block 連接的方式，通過逐層將該層與上一層的注意力矩陣進行矩陣相乘，來模擬特徵的傳播路徑，如式(13)所示。為了考慮 self-attention layer 間的 residual connections，因此在相乘前會讓注意力矩陣加上一個單位矩陣，並進行 normalize，如式(14)。最後計算出注意力從輸入至輸出的聚焦的位置。其中$\tilde{A}$ 代表該層經過 Attention Rollout 後的結果，A 代表原始的注意力矩陣。l 則是 attention layer 的層數，在輸入層則將 j 設為 0。W_{attn} 代表了注意力權重。

$$\tilde{A}(l_i) = \begin{cases} A(l_i)\tilde{A}(l_{i-1}), & if\ i > j \\ A(l_i), & if\ i = j \end{cases} \quad (13)$$

$$A = 0.5\,W_{attn} + 0.5\,I \quad (14)$$

4.2.6 性能指標

為判斷分類模型預測的能力，常用的效能指標有正確率（Accuracy）、靈敏度（Sensitivity）、特異度（Specificity）與 AUC（Area under curve），這些指標可透過混淆矩陣（confusion matrix）在呈現分類結果的四個象限來計算得出。混淆矩陣（圖 4.10）中對照本實驗的四種分類情形分別為：

1. True positive (TP)：實際上為 PD-L1≥50%(+)被預測為 PD-L1≥50% (+)
2. True negative (TN)：實際上為 PD-L1<50% (−)被預測為 PD-L1<50% (−)
3. False positive (FP)：實際上為 PD-L1<50% (−)被預測為 PD-L1≥50% (+)
4. False negative (FN)：實際上為 PD-L1≥50% (+)被預測為 PD-L1<50% (−)

計算出以上四種分類情形後，即可計算出上述分類指標 Accuracy、Sensitivity、Specificity 以及 Area Under Curve（AUC）。

	Predicted Positive	**Predicted Negative**
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

圖 4.10 confusion matrix

準確度（Accuracy）指對於真實情況的預測能力，代表本研究的分類模型正確分類出 PD-L1≥50%(+)與 PD-L1<50% (−)的比例，如式（ 15)所示。

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \qquad (15)$$

靈敏度（Sensitivity）指對於真實情況中 positive class 的預測能力，表示本研究之分類模型正確分辨出 PD-L1≥50% (+)的比例，如式(16)所示。

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \qquad (16)$$

特異性（Specificity）指對於真實情況中 negative class 的預測能力，表示本研究之分類模型正確分辨出 PD-L1<50% (−)的比例，如式(17)所示。

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \qquad (17)$$

AUC 是指在 ROC 曲線（Receiver Operating Characteristic）下的面積。ROC 曲線是一種用於衡量分類器性能的二維圖形，其中橫坐標為假陽性率（False Positive Rate），常用 1 - Specificity 表示；縱坐標為真陽性率（True Positive Rate），也稱做 Sensitivity。其特性是能夠清楚地展示出分類器的效果，並比較不同分類器之間的性能。ROC 曲線的計算方法為：對於每個預測結果，計算 False Positive Rate 和 True Positive Rate，並將其連接形成曲線。AUC 將 ROC 曲線下的面積求和，其值範圍為 0 到 1，AUC 越大，表示分類效能越好。AUC＝1 時，表示模型擁有最佳的預測能力；AUC＞0.5 時，表示模型相較隨機猜測來的好，更具判斷能力；AUC ≤0.5 時，則表示模型預測結果與隨機猜測相同或更差，不具有參考價值。

第五章 研究結果與討論

5.1 MTMAE 分類結果

由於醫學影像和自然影像有極大的差異，不論在解析度、對比度、亮度、以及影像的認知和解讀方式上都不相同。為了在 transfer learning 上能供足夠的醫學影像作為 pre-train 資料，避免使用自然影像產生的 domain 轉換問題。因此，本研究在預訓練資料中採用 GAN 模型生成之肺結節影像，GAN 模型使用 LIDC-IDRI 開放資料集中 1375 筆資料做訓練。GAN 取用 LIDC-IDRI 資料集中平均惡性腫瘤分數 ≤ 2 以及 ≥ 4 的樣本，並以之分為良性及惡性結節，前者 823 筆後者則是 462 筆。GAN 模型在搭配不同肺實質背景與二元結節標記區域後產生生成樣本，為了避免腫瘤過小而喪失影像上的顯著性，本研究限制腫瘤體積需大於 3300 mm^3，最後得到 40,618 筆 3D 生成樣本，並間隔取出 518,064 張 2D 影像。影像前處理的腫瘤區域判定部分，在 LIDC-IDRI 資料集中有四位放射科醫師所標記之腫瘤邊界，以四位醫師標記 50%的共同區域作為腫瘤區域的範圍。

MTMAE 在預訓練任務中採用 multi-task learning，分別執行了重建以及分割兩個任務，幫助模型針對醫學影像特性訓練。圖 5.1 為使用 PD-L1 資料做為測試影像的結果（上三張為 PD-L1 positive，下三張為 PD-L1 negative 影像）。圖 5.1 中從左到右分別是原圖、經過 75% masking 後進入 encoder 的影像、重建後的影像、腫瘤邊界的二元 mask 答案，以及 model 預測的腫瘤邊界。

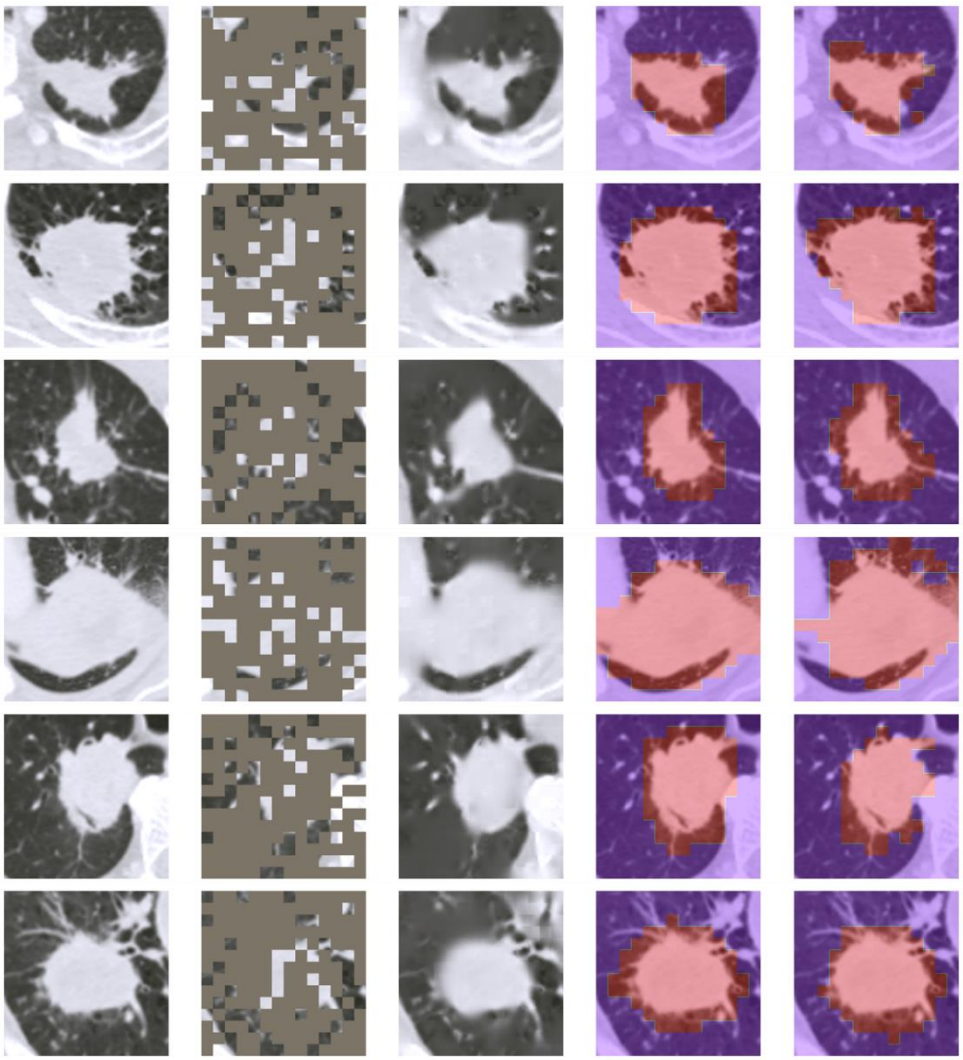

圖 5.1 Pre-train 結果，從左到右分別是 original image、masked image、reconstructed image、original tumor boundary，以及 predicted boundary。（紅色為腫瘤區域，藍紫色為背景區域）

Fine-tuning PD-L1 分類的部分，本研究選採用 3-fold stratified cross validation 的交叉驗證方法，將原始資料集隨機切分為 3 等份，每份都根據資料中的類別比例抽樣得到，以保證每個 fold 中的資料類別分佈相似。接著，我們會輪流選擇其中一份作為測試集 （testing data），而剩下的資料中的 3/4 會被用來當作訓練集（training data），另外 1/4 則會被用來當作驗證集（validation data）。輪流選擇不同的 fold 當作測試集，可以保證每個 fold 都會成為測試集，從而有效地避免了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。

本研究總共蒐集 188 例病患，其中 PD-L1 expression ≥ 50%(+)有 49 例，PD-L1 expression < 50% (−)有 139 例。經過 data augmentation 後每個 fold 平均為 3295 筆 training data，31 筆 validation data 以及 62 筆 testing data，其資料分布如表 5.1 所示，表 5.2 為三次 3-fold cross validation 結果之平均。經過三次結果平均，AUC 為 0.735、準確率為 0.724、靈敏度為 0.583，特異度為 0.773。

表 5.1 資料集分配表

	PD-L1 ≥ 50%(+)	**PD-L1 < 50% (−)**	**Total**
Training data (2D images, ave)	25 (1655)	70 (1640)	95 (3295)
Validation data	8	23	31
Testing data	16	46	62

表 5.2 三次 3-fold cross validation 之平均

Round	**AUC**	**Accuracy**	**Sensitivity**	**Specificity**
1	0.744	0.721	0.729	0.717
2	0.731	0.720	0.541	0.783
3	0.731	0.731	0.479	0.819
Average (±σ)	0.735 ± 0.003	0.724 ± 0.002	0.583 ± 0.043	0.773 ± 0.017

5.2 消融實驗

為了探究不同因素對本研究提出之模型的影響，本研究進行了一系列的消融實驗，涵蓋了三個關鍵方面：預訓練方法、預訓練任務和預訓練資料集大小。這些實驗旨在研究這些因素與本研究提出之模型特點之間的關係。

首先，採用自監督學習作為預訓練方法，以增強模型的遷移能力並降低有效訓練所需的資料門檻。通過利用無標籤資料，模型可以學習到有用的表示，並能更好的適應不同的資料集。其次，於預訓練階段引入了一個分割任務。這個額外的任務旨在增強模型對腫瘤區域的關注，並提高其提取腫瘤特徵的能力。通過訓練模型區分前景（腫瘤）和背景區域，進而在 PD-L1 表達分類中獲得更好的性能。最後，通過使用生成對抗網絡（GAN）來擴增資料量，改變了預訓練資料集的大小。通過生成更多樣化的樣本，旨在研究資料集大小對模型捕捉廣泛的腫瘤特徵和改善其泛化能力的影響。

5.2.1 預訓練方法

本研究的目的在於透過自監督預訓練方法提高模型的遷移能力，從而克服在建立深度模型時受限於資料量的問題。本節將比較本研究提出的 MTMAE 自監督預訓練架構與傳統的監督預訓練方法，遷移至下游任務在 PD-L1 表現分類上的差異。以先前將自監督訓練應用於醫學影像的研究為例[24]–[26]，先前文獻觀察到使用自監督方法進行預訓練的模型在分類醫學影像任務中，取得了優於監督預訓練方法的結果。在本節比較中，我們的方法和傳統監督預訓練方法皆採用了如 5.1 所描述之 GAN 生成影像作為預訓練材料。然而，兩種方法在預訓練任務的設計上有所不同。在自監督預訓練方法中，採用了重建和分割作為預訓練任務，讓模型學習醫學影像的特徵。而在監督預訓練方法中，由於 GAN 是基於 LIDC-IDRI 資料集生成的，以分類良惡性結節作為預訓練任務。透過比較這兩種不同的預訓練方法，可以觀察到不同的預訓練方法對下游任務的遷移能力所產生的影響。

實驗結果顯示，使用本研究提出的 MTMAE 自監督預訓練架構方法之平均三次 3-fold cross validation 的 AUC 為 0.735 ± 0.003，平均準確率為 0.724 ± 0.002。相較之下，使用傳統監督預訓練方法之平均 AUC 為 0.695 ± 0.008，平均準確率為

0.676 ± 0.006。本研究所提出的自監督預訓練 MTMAE 架構在 PD-L1 分類上取得了較好的表現，由此可推論自監督學習方法在遷移能力與其對下游任務的泛用性上，較監督預訓練方法更具有潛力。表 5.3 顯示了三次 3-fold cross validation 的在不同預訓練方法下的結果。表 5.4 為使用監督預訓練方法在 PD-L1 分類上的表現，本研究所提出的自監督預訓練 MTMAE 架構在 PD-L1 分類之表現則見表 5.2。

自監督學習方法勝出的原因可能源於其使用了無需人工標記的預訓練任務，從而提供了更廣泛的學習範圍。同時，也讓模型自行學習醫學影像中的隱含的特徵，進而更好地捕捉影像中的結構和特徵。相比之下，監督學習方法依賴於人工標記的訓練數據，可能對與特定標籤相關的特徵產生偏重，導致模型在面對新的、不同分佈的下游任務影像時表現不如預期。此外，標記品質也可能對監督學習方法的結果產生影響。

研究結果與先前在醫學影像中使用自監督方法分類的相關研究一致，顯示在醫學影像中使用自監督預訓練對於建立遮蓋圖像模型是一種有效的策略。這一發現有助於後續在醫學影像領域對深度學習的應用。然而，儘管自監督預訓練方法在本研究中取得了良好的結果，但仍然需要進一步的研究來深入瞭解不同預訓練方法的優缺點及其在不同場景下的適用性。

表 5.3 不同預訓練方法的 PD-L1 分類結果比較

	Self-Supervised pre-training	Supervised pre-training
AUC	0.735 ± 0.003	0.695 ± 0.008
Accuracy	0.724 ± 0.002	0.676 ± 0.006
Sensitivity	0.583 ± 0.043	0.590 ± 0.046
Specificity	0.773 ± 0.017	0.705 ± 0.024

表 5.4 監督預訓練方法於 PD-L1 分類結果三次 3-fold cross validation 之平均

Round	AUC	Accuracy	Sensitivity	Specificity
1	0.675	0.688	0.500	0.753
2	0.688	0.683	0.521	0.739
3	0.721	0.656	0.750	0.623
Average (±σ)	0.695 ± 0.008	0.676 ± 0.006	0.590 ± 0.046	0.705 ± 0.024

5.2.2 預訓練任務

Masked Autoencoder (MAE) 架構在自然影像上取得了良好的結果，然而，在醫學影像中，目標物往往缺乏自然影像中的突出和明顯特徵，並容易與背景混合。這使得僅進行重建預訓練任務的 MAE 模型難以準確地捕捉到醫學影像中目標物的結構和特徵。為了克服這一挑戰，本研究提出了專為醫學影像的特性而設計的 MTMAE 架構。MTMAE 採用了 multi-task learning 的架構，並引入了一個新的 Segmentation Decoder。這一設計使得 MTMAE 模型能夠同時學習重建和分割兩個任務，以更好地理解醫學影像中的資訊並具備區分腫瘤內外的能力。醫學影像中的目標物常常與周圍組織和背景混合，因此，透過分割任務的訓練，MTMAE 模型能夠更加專注於區分前景和背景，並學習捕捉腫瘤區域的特徵。

為了深入瞭解 MTMAE 模型在應對醫學影像挑戰方面的優勢，本節實驗了使用原始 MAE 架構僅進行重建任務（Reconstruction Decoder）的預訓練模型，以及使用 MTMAE 架構的訓練重建和分割兩個任務的預訓練模型（Reconstruction Decoder + Segmentation Decoder），並觀察兩者在遷移至下游 finetune 用於分類 PD-L1 表現的任務上的差異。

在比較結果中，使用 MTMAE 模型進行預訓練之平均三次 3-fold cross

validation 的 AUC 為 0.735 ± 0.003，平均準確率為 0.724 ± 0.002。相比之下，僅使用重建任務的原始 MAE 模型的平均 AUC 為 0.712 ± 0.002，平均準確率為 0.735 ± 0.004。這表明在預訓練階段同時訓練重建和分割兩個任務的 MTMAE 模型在下游任務的表現上優於僅進行重建任務的 MAE 模型。表 5.5 為兩種模型經過三次 3-fold cross validation 的結果比較，表 5.6 為 MAE 架構僅訓練重建任務在 PD-L1 分類之表現，MTMAE 架構訓練重建和分割兩個任務模型在 PD-L1 分類之表現見表 5.2。

此外，本實驗利用 4.2.5 所提及之 Attention Rollout 方法對模型的關注區域進行視覺化，以幫助理解模型學習到的特徵。圖 5.2 展示了 MTMAE 與 MAE 兩個模型的 attention score 視覺化結果。由上面兩列可得知，當遇到與背景顏色相近的早期腫瘤時，由於尺寸較小且與背景混合，這些區域較難被單純進行重建任務的 MAE 模型找到。然而，由於 MTMAE 模型在預訓練階段同時學習了重建和分割任務，它對腫瘤前景和背景有較好的認知，能夠更清晰地偵測到腫瘤並進行學習。觀察下面兩列，當腫瘤位置明顯時，MTMAE 模型相對於單一任務的 MAE 模型對腫瘤主體仍具有更高的關注度，且對背景其他組織（如肺壁、心臟）的影像程度較低。透過加入 Segmentation Decoder 並訓練分割任務，模型被迫專注於區分前景和背景，從而學習到更具區分度的特徵。這使得 MTMAE 模型能夠更好地理解醫學影像中目標物的結構和特徵，尤其在目標物與周圍組織和背景混合的情況下。

MTMAE 模型相對於僅使用重建任務的 MAE 模型具有明顯的優勢，主要原因在於它同時學習了重建和分割兩個任務，從而獲得更全面和具有判別性的影像特徵。這使得模型能夠更好地處理醫學影像中目標物的結構和特徵，並提高下游任務的表現。

表 5.5 不同預訓練任務的 PD-L1 分類結果比較

	Multi-task (MTMAE)	Single-task (MAE)
AUC	0.735 ± 0.003	0.712 ± 0.002
Accuracy	0.724 ± 0.002	0.735 ± 0.004
Sensitivity	0.583 ± 0.043	0.458 ± 0.025
Specificity	0.773 ± 0.017	0.831 ± 0.013

表 5.6 原始 MAE 架構於 PD-L1 分類結果三次 3-fold cross validation 之平均

Round	**AUC**	**Accuracy**	**Sensitivity**	**Specificity**
1	0.715	0.737	0.521	0.812
2	0.706	0.747	0.375	0.877
3	0.716	0.721	0.479	0.805
Average (±σ)	0.712 ± 0.002	0.735 ± 0.004	0.458 ± 0.025	0.831 ± 0.013

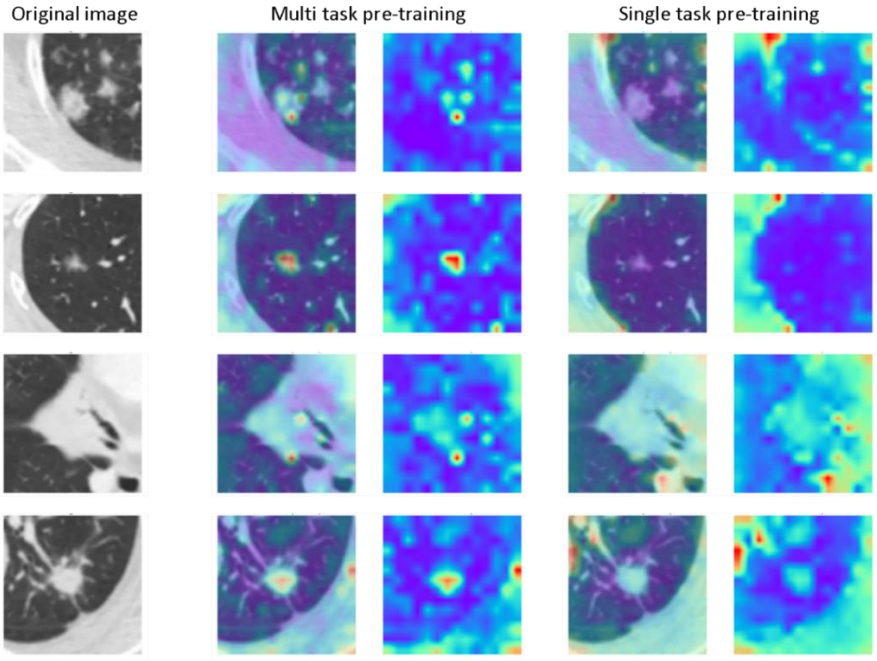

圖 5.2 不同預訓練任務的 Attention map

5.2.3 預訓練資料集大小

MTMAE 作為一個 Transformer-based 模型，具有大量的參數。在訓練過程中，資料量的多寡對於其學習的效果有很大的影響。在醫學領域，由於資料的稀缺性和隱私問題，蒐集足夠數量和多樣性的真實醫學影像資料是一項具有挑戰性的任務。為了克服這些困難，本研究採用生成對抗網路（GAN）作為解決方案，GAN 具有生成大量一般蒐集不到的資料的能力，且生成之影像具有豐富的多樣性，能夠填補真實資料量的不足。另外，由於 GAN 使用腫瘤邊界作為生成輸入，使用 GAN 生成之影像作為預訓練資料來源即可直接獲得已知腫瘤邊界，不需擔心邊界答案來源問題。在本研究中，我們運用 GAN 模型來擴充預訓練資料集的大小和多樣性，從而幫助 MTMAE 更好地學習資料的特徵。

本研究使用之 GAN 模型在模擬真實結節的紋理下增加結節紋理特徵之多樣性，同時也考慮到結節形狀與肺實質背景之多樣性。本節的目標是比較使用本研究之 GAN 模型生成的大量資料和僅使用一般能蒐集到之真實醫學影像資料集兩者之間的差異（資料列於表 5.7）。我們分別使用這兩種資料集進行預訓練，然後進行下游的 PD-L1 表現分類，並比較它們的表現，以驗證在本實驗中使用 GAN 的影響和幫助。為了減少其他變因的影響，此處的預訓練任務使用原始 MAE 模型的遮蓋還原。

在真實資料組的資料集中，除了使用了訓練 GAN 的 LIDC-IDRI 公開資料集外，還額外加入了臺大醫院的 CT 影像以增加資料量。臺大醫院的影像經過實驗室先前開發的影像分割演算法進行分割，再由醫師進行確認。真實資料組由 1476 筆 LIDC-IDRI 公開資料集和 530 筆臺大醫院的 3D CT 影像組成，遵循 GAN 篩選條件，僅使用腫瘤體積大於 3300 mm^3 的樣本，經過前處理後總共得到 13,717 筆 2D 影像。

實驗結果顯示，使用 GAN 的模型平均三次 3-fold cross validation 的 AUC 為 0.712±0.002，平均準確率為 0.735±0.004。使用真實資料的模型平均 AUC 為 0.704 ±0.006，平均準確率為 0.708±0.01。相較於未使用 GAN 的模型，使用 GAN 生成的資料的模型表現更好，這表明當模型擁有更多樣本時，有助於學習更具泛化性和有效性的特徵。使用 GAN 模型作為資料增量的方法能夠克服深度學習中學習能力受限而導致泛化能力不佳的問題，同時也提供了更好的遷移能力。GAN 模型生成的擬真樣本能夠為模型提供更多的多樣性，這有助於增強模型對於各種不同情況和變化的適應能力。表 5.8 為三次 3-fold cross validation 在不同資料量的模型之結果比較。表 5.9 為使用真實資料預訓練並遷移至 PD-L1 分類任務時的表現，表 5.6 則展示了使用 GAN 模型進行資料增量的資料集預訓練後遷移至 PD-L1 分類任務時的表現。

表 5.7 兩種預訓練資料集的數量

Data	GAN	真實資料
CT 樣本數量	40,618	2,006
訓練 2D 影像數量	518,064	13,717

表 5.8 不同資料集大小對 PD-L1 分類結果的比較

Data	GAN	真實資料
AUC	0.712 ± 0.002	0.704 ± 0.006
Accuracy	0.735 ± 0.004	0.708 ± 0.01
Sensitivity	0.458 ± 0.025	0.597 ± 0.022
Specificity	0.831 ± 0.013	0.746 ± 0.021

表 5.9 使用真實資料集大小 pre-train 在 PD-L1 分類之表現

Round	AUC	Accuracy	Sensitivity	Specificity
1	0.703	0.677	0.646	0.689
2	0.721	0.737	0.521	0.811
3	0.687	0.710	0.625	0.739
Average (±σ)	0.704 ± 0.006	0.708 ± 0.01	0.597 ± 0.022	0.746 ± 0.021

另外，在醫學影像任務中，常會使用 ImageNet 進行預訓練，ImageNet 為資料量多達 1400 萬筆的自然影像資料集。先前的文獻中指出[25]，使用 ImageNet 進行預訓練可以取得較好的結果，只在特定實驗資料集中有明顯差異。因此，此節另外比較本研究使用 GAN 生成的醫學影像和 ImageNet 預訓練的模型在下游 PD-L1 分類任務中的表現。與上一實驗相同，為了減少其他變因的影響，預訓練任務使用原始 MAE 模型的遮蓋還原。

實驗結果顯示，使用 GAN 的模型平均三次 3-fold cross validation 的 AUC 為 0.712 ± 0.002，平均準確率為 0.735 ± 0.004。使用 ImageNet 的模型平均 AUC 為 0.669 ± 0.009，平均準確率為 0.687 ± 0.009。相較於使用 ImageNet 的模型，使用 GAN 生成的資料的模型表現更優越。表 5.10 為三次 3-fold cross validation 在不同資料量的模型之結果比較。表 5.11 為使用 ImageNet 預訓練並遷移至 PD-L1 分類任務時的表現，表 5.6 則展示了使用 GAN 模型進行資料增量的資料集預訓練後遷移至 PD-L1 分類任務時的表現。

在 PD-L1 分類任務中，使用 GAN 生成的資料能夠提供更具代表性的樣本，幫助模型更好地適應各種情景和變化。這與 Xu[25]所得到的結果相反，可能是因為其任務為分辨 COVID CT 影像，而該任務與自然影像較相近，能用肉眼輕易分辨。然而，PD-L1 表現量這樣缺乏明顯外觀特徵的任務，使用 ImageNet 的自然影像進

行訓練將會產生極大的 domain gap。結果顯示，使用 GAN 模型作為資料增量的方法能夠直接使用相同領域的影像訓練，並更有效地提取細微的特徵，從而改善模型的性能。

表 5.10 不同資料來源對 PD-L1 分類結果的比較

Data	GAN	ImageNet
AUC	0.712 ± 0.002	0.669 ± 0.009
Accuracy	0.735 ± 0.004	0.687 ± 0.009
Sensitivity	0.458 ± 0.025	0.458 ± 0.028
Specificity	0.831 ± 0.013	0.766 ± 0.022

表 5.11 使用 ImageNet pre-train 在 PD-L1 分類之表現

Round	AUC	Accuracy	Sensitivity	Specificity
1	0.657	0.683	0.458	0.761
2	0.700	0.715	0.375	0.834
3	0.649	0.661	0.542	0.703
Average (±σ)	0.669 ± 0.009	0.687 ± 0.009	0.458 ± 0.028	0.766 ± 0.022

透過以上兩個實驗可以發現，引入 GAN 生成的資料，表現顯著提升，證明了 GAN 模型作為資料增量的有效性。此外，結果發現預訓練之影像來源對模型的特徵提取有極大的影響。特別是在需要精細特徵的任務中，如 PD-L1 表現量預測，使用自然影像預訓練的效果並不理想。進一步說明了在訓練大型模型時，資料量和資料來源的選擇都是重要因素。通過使用 GAN 生成額外的資料，能夠合成更多的醫學影像樣本，從而擴大資料集直接使用醫學影像訓練。這使得模型能夠更好地捕捉資料分佈中的細微變化和特徵，提高對於不同情境的泛化能力，增強了模型的穩

定性和適用性。

5.3 重現文獻之方法

除了比較上述方法，也嘗試使用本研究蒐集之資料重現第二章中近年來文獻所提出的架構進行比較。本節針對與本實驗同樣使用深度學習特徵之文獻做重現比較。其中，Zhu[19]單純使用深度學習特徵，Tian[20]和 Wang[21]兩篇文獻則是結合放射體學、深度學習與臨床特徵。三篇文獻中，僅有 Zhu 使用 5-fold 交叉驗證，其餘兩篇 Tian 隨機挑出 1/10 的病患，Wang 則是 1/5 的病患做為測試集。然而，為了與本實驗結果比較以及得到較穩定的表現，表 5.12 本研究資料重現文獻之結果是重現各方法並且重複三次 3-fold stratified cross validation 之結果。

在深度學習的部分，Zhu[19]等人使用修改 2D DenseNet 的架構成 3D 形式，以 3D DenseNet 來提取影像資訊，並且使用 ImageNet 進行 transfer learning。該文獻對 127 位晚期肺腺癌患者進行 PD-L1 表現量≥50%的分類中，其 AUC 為 0.765。本研究蒐集資料集重現在 Zhu 之方法上，AUC 為 0.588，準確率為 0.613。Tian[20]等人先將影像變成許多 2D 影像，再放入 2D DenseNet 做影像提取。該文獻共採納 939 位晚期病人，單獨使用深度學習方法的 AUC 為 0.68。本研究蒐集資料集重現在 Tian 之方法上，單獨使用深度學習方法 AUC 為 0.642，準確率為 0.669。Wang[21]在深度特徵上使用 3D ResNet 提取，並減少 subsampling 來配合較小的輸入影像。該文獻將 PD-L1 表現量分成高（≥50%）、中（1%-49%）、低（<1%）三種，在病人數 1135 位分類 PD-L1 表現量≥50%的實驗中，該文獻單獨使用深度學習方法的 AUC 為 0.901。由於該文獻中提及有使用較大的資料集進行預訓練，因此在重現時使用 LIDC 資料集中分辨結節良惡性之任務進行預訓練。本研究蒐集資料集重現在 Wang 之方法上，單獨使用深度學習方法 AUC 為 0.608，準確率為 0.701。由於該文獻並無給予詳細模型架構，因此嘗試三種不同大小之 ResNet，並按照文獻中減

少 subsampling 之修改，得到 3D ResNet18 之 AUC 為 0.58，3D ResNet34 之 AUC 為 0.608，3D ResNet50 之 AUC 為 0.468，選擇其中表現最好的 ResNet34。

基於 Tian[20]的研究中單獨使用放射體學特徵的表現良好，本研究也與 Tian[20]和 Wang[21]使用之放射體學方法進行重現比較。這兩篇文獻並未提供放射體學的分類器或作為分類器的全連階層的參數，因此在以下的實驗中，使用了常見的分類器，如 Logistic regression、SVM、Random forest 和單層的全連階層，並選擇表現最佳的分類器作為結果。Tian[20]套入不同 filter 來提取 radiomics 特徵，再經由 Mann Whitney U test 挑選特徵。該文獻單獨使用放射體學方法的 AUC 為 0.75。本研究蒐集資料集重現在 Tian 之方法上，單獨使用放射體學方法 AUC 為 0.551，準確率為 0.59。Wang[21] 提取的 radiomics 特徵包括形狀、一階及二階等紋理特徵，並以 Lasso 與 variance 作為篩選機制。在分類 PD-L1 表現量≥50%的實驗中，單獨使用放射體學方法的 AUC 為 0.946。本研究蒐集資料集重現在 Wang 之方法上，單獨使用放射體學方法 AUC 為 0.544，準確率為 0.614。

表 5.12 本研究資料重現文獻之結果

Method	Model	Data	DL AUC	ML AUC
Zhu[19]	3D DenseNet	文獻	0.765	-
		本研究	0.588	-
Tian[20]	2D DenseNet	文獻	0.68	0.75
		本研究	0.642	0.551
Wang[21]	3D ResNet	文獻	0.901	0.88
		本研究	0.608	0.544

另外，本研究亦統整各文獻中挑選出有鑑別力的 radiomics 特徵，回顧的六篇中的有五篇包含 radiomics 特徵，其中有四篇有列出所使用的特徵。表 5.13 為各文

獻挑選有鑑別力的 radiomics 特徵在本研究蒐集資料集中之 p-value，使用與文獻相同的 Mann Whitney U test。在進行重現時，使用本實驗的資料來重現原文獻的方法，以確保結果的一致性。然而，實驗結果顯示，無論是使用深度學習方法或是放射體學方法，我們皆無法達到與原文獻相同的表現。在深度學習方法方面，由於本實驗樣本數較少，即使於重現之三種深度學習方法皆有經過預訓練，但訓練一般的深度學習網路架構仍然較困難。在醫學影像領域，資料集通常規模較小，例如重現方法中最大的資料集約為一千筆樣本。相較於自然影像上至千萬的資料量，這樣的資料規模對於建立和驗證複雜的深度學習模型來說是有限的，容易過擬合於訓練樣本或導致學習特徵不具泛用性。因此，文獻中的方法尚需要進一步的驗證，以確定在其他資料集上的通用性和醫學實際應用的效果，否則我們無法確定這些方法是否適用於我們的資料集。

放射體學方法上，除了重現兩篇文獻之方法，也針對四篇文獻中所使用之特徵進行比較。雖然該四篇文獻於原文獻之資料皆有良好的結果，但這些文獻所選取的特徵集合並不存在重疊，且在我們的資料集上進行統計分析時，所得到的 p-value 偏高，不具有統計學上的顯著區分能力。結果顯示這些 radiomics 方法未能歸納出一個能夠顯著分辨 PD-L1 的特徵，同時突顯了這些方法的泛用性有限，在未來的研究中，仍需要進一步探索。

對於重現文獻方法之結果差異，推測可能來自多個因素的綜合影響。首先，資料分布的不同可能是一個重要因素。不同的研究使用來自不同機構、不同地理區域的資料集，使用的 CT 設備設定也不同，這些資料集在病人樣本和腫瘤特徵上存在差異。除了資料集的分布差異，還有其他因素可能導致我們的重現結果與原文獻不一致。例如，在免疫組織化學染色中，使用不同的染色抗體可能會導致結果的差異。各個抗體之間的 PD-L1 表現一致性不同，因此也可能會導致各研究提取到不同特徵。除此之外，文獻之間對採納病人的差異也是變因之一，像是病人的分期和病理

分類也可能對 PD-L1 預測結果產生間接的影響。這些因素可能與腫瘤的發展和進展相關，因此可能會對 PD-L1 的表達模式產生變化。不同分期和病理分類的病人可能具有不同的臨床特徵和腫瘤特性，這可能會影響到 PD-L1 的表現和分佈。

由於本實驗樣本數較少，無法作為一個公正的基準來比較各預測方法的表現，因此在討論結果時謹慎地提出以上可能導致重現不如預期的原因。因為資料量不足，可能導致穩定性與泛用性低的問題，這也是醫學影像領域在資料收集和方法驗證方面所面臨的挑戰。然而，實驗結果仍顯示出在如本研究的小型資料集中，本研究提出的 MTMAE 架構相較於傳統的監督式預訓練搭配 CNN 模型表現更為突出。這也強調了在研究設計和資料收集階段的重要性，除了設計克服資料量依賴問題的模型，未來的研究應進一步擴大資料集的規模，確保所提出的方法在不同資料群體上的穩健性，從而提高預測模型的可靠度。

表 5.13 文獻挑選出之 radiomics 特徵於本研究樣本的 p-value

Method	Features	p-value	
		文獻樣本	本研究樣本
Jiang[16]	Wavelet1-LLH-GLSZM- large area high gray level emphasis	<0.05	0.592
	Wavelet1-LHL-interquartile range	<0.05	0.756
	Wavelet1-HLH-NGTDM-busyness	<0.05	0.752
	Wavelet2-HHL-GLSZM-gray level non-uniformity	<0.05	0.939
	Wavelet2-HHH-GLSZM-zone entropy	<0.05	0.312
	Wavelet2-HHH-GLDM-dependence entropy	<0.05	0.185
	Wavelet2-HHH-GLDM-dependence entropy variance	<0.05	0.490
	Wavelet2-LLL-GLSZM-large area high gray level emphasis	<0.05	0.779
	LBP3D-10 percentile	<0.05	0.914
	maximum	<0.05	0.311
Bracci[17]	skewness	<0.01	0.972
	GLZLM-low gray level zone emphasis	0.049	0.970
Wen[18]	kurtosis	0.033	0.948
	GLCM-cluster tendency	0.005	0.710
	GLSZM-size zone non-uniformity	0.012	0.927
	GLRLM-gray level non-uniformity normalized	<0.001	0.749
	Wavelet1-HLH-GLRLM-long run high gray level emphasis	0.006	0.772
	Wavelet1-HLL-GLSZM-high gray level zone emphasis	<0.001	0.455
Tian[20]	Wavelet1-HL-GLCM-sum entropy (2D)	-	0.937
	Wavelet1-LL-GLRLM-gray level non-uniformity normalized (2D)	-	0.754

第六章 結論與未來展望

肺癌已成為全球最常見的癌症之一，而針對肺癌的治療策略一直是臨床上的重要議題。比起過去缺乏專一性的治療，近年來 NSCLC 的治療研究也更專注在具專一性的治療策略，針對腫瘤細胞進行控制。PD-1/PD-L1 抑制劑已顯示了在晚期 NSCLC 治療上有顯著效果。其中，腫瘤的 PD-L1 expression 已成為目前採用的 biomarker，也是決定是否適合免疫治療的關鍵之一。然而，現有的評估方式存在限制，包括檢體難以反應整體表現或判定標準缺乏共識等問題，導致準確率不穩定。為了克服現有檢測的不足，本研究希望利用非侵入性且能夠整體判讀腫瘤的 CT 影像來建立使用深度模型的電腦輔助診斷系統（CAD）。然而，醫學影像樣本用於訓練深度模型相對不足。

為了克服資料量問題，本研究提出 Multi-task Masked Autoencoder（MTMAE）模型。其中，使用自監督學習架構中遮蓋圖像模型來提高模型的遷移能力，並且降低資料的門檻。同時，使用 GAN 完成資料增量，生成的大量醫學資料訓練，使模型能夠學習到多樣豐富的特徵。另外，為了使遮蓋圖像模型更適應醫學影像特性，避免目標物不明顯的問題，在多任務學習中加入分割任務，使模型在提取特徵時能夠區分前景和背景，更好地捕捉腫瘤的特徵。MTMAE 在本研究資料中，應用於 PD-L1 50%表現量分類 AUC 為 0.735，準確率為 0.724。在消融實驗中，相比於傳統的監督式預訓練和訓練單一重建任務的 MAE 表現更好；使用 GAN 生成大量醫學樣本的表現亦勝過使用 ImageNet 與未使用生成資料的。從中可以更好地理解本研究提出之方法的優勢。

由實驗結果可看出自監督訓練在醫學影像中的潛力，以及將其結合 GAN 的連帶益處。然而，由於僅使用了一個 PD-L1 資料集且資料量有限，未來需要進一步增加資料量並進行更多實驗來驗證相關結果。在 PD-L1 影像的電腦輔助診斷系統

方面，儘管許多文獻已經取得不錯的成果，但我們也觀察到這些成果與資料的分佈有一定程度的相關性，難以重複。資料分布同時也是資料量不足所引起的取樣偏差，如何克服資料量限制問題，這仍然是醫學影像領域待解的重要議題。未來若能發展多機構、多區域的合作研究，將能提供更廣泛的資料集和更具代表性的樣本，從而增強模型的泛化能力和可靠性。

基於資料量的限制，本研究發展的 PD-L1 表現量預測模型使用了二維 CT 影像採樣。未來，若能夠蒐集到足夠的資料，可以發展三維模型架構，以更充分地利用 CT 影像的特徵。此外，在幫助篩選適合使用 PD-L1 免疫治療的患者方面，僅使用影像預測 PD-L1 表現量只是第一步。目前也存在其他與治療預後有關的因子，如腫瘤突變負荷和腫瘤浸潤淋巴細胞，雖然這些因子的研究尚未成熟，但也是後續實驗可以考慮的資料。為了克服目前 PD-L1 表現量檢測上的問題，最終目標是讓影像生物標誌能夠進一步預測病人的治療結果。若能蒐集到足夠的病人治療結果資料，將能與現階段的研究結合，從而提升預測模型的準確性和實用性，將增加在臨床上的應用價值。

參考文獻

[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA: A Cancer Journal for Clinicians, vol. 68, no. 6, pp. 394–424, 2018.

[2] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, "Cancer statistics, 2022," CA: A Cancer Journal for Clinicians, vol. 72, no. 1, pp. 7–33, 2022.

[3] D. Planchard et al., "Metastatic non-small cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†," Annals of Oncology, vol. 29, pp. iv192–iv237, 2018.

[4] M. Reck et al., "Pembrolizumab versus Chemotherapy for PD-L1–Positive Non–Small-Cell Lung Cancer," New England Journal of Medicine, vol. 375, no. 19, pp. 1823–1833, 2016.

[5] R. S. Herbst et al., "Atezolizumab for First-Line Treatment of PD-L1–Selected Patients with NSCLC," New England Journal of Medicine, vol. 383, no. 14, pp. 1328–1339, 2020.

[6] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." arXiv, 2019.

[7] Y. Ishida, Y. Agata, K. Shibahara, and T. Honjo, "Induced expression of PD-1, a novel member of the immunoglobulin gene superfamily, upon programmed cell death.," The EMBO Journal, vol. 11, no. 11, pp. 3887–3895, Nov. 1992.

[8] H. Nishimura, M. Nose, H. Hiai, N. Minato, and T. Honjo, "Development of Lupus-like Autoimmune Diseases by Disruption of the PD-1 Gene Encoding an ITIM Motif-Carrying Immunoreceptor," Immunity, vol. 11, no. 2, pp. 141–151, 1999.

[9] G. J. Freeman et al., "Engagement of the Pd-1 Immunoinhibitory Receptor by a Novel B7 Family Member Leads to Negative Regulation of Lymphocyte Activation," Journal of Experimental Medicine, vol. 192, no. 7, pp. 1027–1034, 2000.

[10] H. Dong et al., "Tumor-associated B7-H1 promotes T-cell apoptosis: A potential mechanism of immune evasion," Nat Med, vol. 8, no. 8, Art. no. 8, 2002.

[11] G. L. Banna et al., "Are anti-PD1 and anti-PD-L1 alike? The non-small-cell lung cancer paradigm," Oncol Rev, vol. 14, no. 2, p. 490, 2020.

[12] R. Brody et al., “PD-L1 expression in advanced NSCLC: Insights into risk stratification and treatment selection from a systematic literature review,” Lung Cancer, vol. 112, pp. 200–215, 2017.

[13] A. N. Niemeijer et al., “Whole body PD-1 and PD-L1 positron emission tomography in patients with non-small-cell lung cancer,” Nat Commun, vol. 9, no. 1, Art. no. 1, 2018.

[14] M. Mathew, R. A. Safyan, and C. A. Shu, “PD-L1 as a biomarker in NSCLC: challenges and future directions,” Ann Transl Med, vol. 5, no. 18, p. 375, 2017.

[15] R. Sun et al., “A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study,” The Lancet Oncology, vol. 19, no. 9, pp. 1180–1191, 2018.

[16] M. Jiang et al., “Assessing PD-L1 Expression Level by Radiomic Features From PET/CT in Nonsmall Cell Lung Cancer Patients: An Initial Result,” Academic Radiology, vol. 27, no. 2, pp. 171–179, 2020.

[17] S. Bracci et al., “Quantitative CT texture analysis in predicting PD-L1 expression in locally advanced or metastatic NSCLC patients,” Radiol med, vol. 126, no. 11, pp. 1425–1433, 2021.

[18] Q. Wen, Z. Yang, H. Dai, A. Feng, and Q. Li, “Radiomics Study for Predicting the Expression of PD-L1 and Tumor Mutation Burden in Non-Small Cell Lung Cancer Based on CT Images and Clinicopathological Features,” Frontiers in Oncology, vol. 11, 2021.

[19] Y. Zhu et al., “A CT-derived deep neural network predicts for programmed death ligand-1 expression status in advanced lung adenocarcinomas,” Ann Transl Med, vol. 8, no. 15, p. 930, 2020.

[20] P. Tian et al., “Assessing PD-L1 expression in non-small cell lung cancer and predicting responses to immune checkpoint inhibitors using deep learning on computed tomography images,” Theranostics, vol. 11, no. 5, pp. 2098–2107, 2021.

[21] C. Wang et al., “Non-Invasive Measurement Using Deep Learning Algorithm Based on Multi-Source Features Fusion to Predict PD-L1 Expression and Survival in NSCLC,” Front Immunol, vol. 13, p. 828560, 2022.

[22] H. Bao, L. Dong, S. Piao, and F. Wei, “BEiT: BERT Pre-Training of Image Transformers.” arXiv, 2022.

[23] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, “Masked Autoencoders Are Scalable Vision Learners,” presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 16000–16009.

[24] L. Zhou, H. Liu, J. Bae, J. He, D. Samaras, and P. Prasanna, “Self Pre-training with Masked Autoencoders for Medical Image Analysis.” arXiv, 2022.

[25] J. Xu and S. Stirenko, “Self-supervised Model Based on Masked Autoencoders Advance CT Scans Classification,” IJIGSP, vol. 14, no. 5, pp. 1–9, 2022.

[26] H. Quan et al., “Global Contrast Masked Autoencoders Are Powerful Pathological Representation Learners.” arXiv, 2022.

[27] A. Vaswani et al., “Attention is All you Need,” in Advances in Neural Information Processing Systems, Curran Associates, Inc., 2017.

[28] A. Dosovitskiy et al., “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale.” arXiv, 2021.

[29] I. Tsimafeyeu et al., “Agreement between PDL1 immunohistochemistry assays and polymerase chain reaction in non-small cell lung cancer: CLOVER comparison study,” Sci Rep, vol. 10, no. 1, Art. no. 1, 2020.

[30] Z. Xie et al., “SimMIM: A Simple Framework for Masked Image Modeling.” arXiv, 2022.

[31] 黃瑋傑, “以生成對抗網路生成電腦斷層掃描三維肺結節樣本： 基於 Gabor 函數之紋理相似性量度與模型選擇指標,” 國立臺灣大學, 2021.

[32] 陳稜鎔, “電腦斷層掃描肺腫瘤良惡性判別之深度學習影像特徵擷取,” 國立臺灣大學, 2018.

[33] S. Abnar and W. Zuidema, “Quantifying Attention Flow in Transformers.” arXiv, 2020.