

肺部電腦斷層掃描之非小細胞癌 PD-L1 表現預測：

結合多任務自監督學習與生成對抗網路

TAI, WEI HSUAN

August 2025

Outline

- 緒論
- 研究目標
- 研究方法
- 研究結果與討論
- 結論與未來展望

Section 1

緒論

研究動機

- 癌症是全球主要的死亡原因之一，肺癌是最常見的癌症之一。
- 肺癌是台灣癌症死亡率最高的癌症類型。

113 台灣癌症統計資料

113年十大癌症死因						
癌症別	死亡人數 (人)		死亡率 (每十萬人口)		標準化死亡率 (每十萬人口)	
		年增率 (%)		年增率 (%)		年增率 (%)
所有癌症死亡原因	54,032	1.7	230.8	1.4	113.3	-1.9
1.氣管、支氣管和肺癌	10,495	1.4	44.8	1.1	21.2	-2.2
2.肝和肝內膽管癌	7,513	-2.7	32.1	-3.0	15.3	-6.3
3.結腸、直腸和肛門癌	7,007	3.2	29.9	2.9	14.2	-0.4
4.女性乳癌	3,050	2.6	25.7	2.2	13.2	-0.8
5.前列腺(攝護腺)癌	1,897	4.5	16.4	4.4	7.8	1.5
6.口腔癌	3,687	2.1	15.7	1.8	8.5	-1.5
7.胰臟癌	2,970	3.2	12.7	2.9	6.1	-0.9
8.胃癌	2,221	-4.6	9.5	-4.8	4.5	-8.3
9.食道癌	2,076	0.6	8.9	0.3	4.7	-2.4
10.卵巢癌	799	6.4	6.7	5.9	3.6	5.1

Figure: 113 年台灣癌症統計資料

資料來源：衛福部官網

2025 美國癌症統計資料

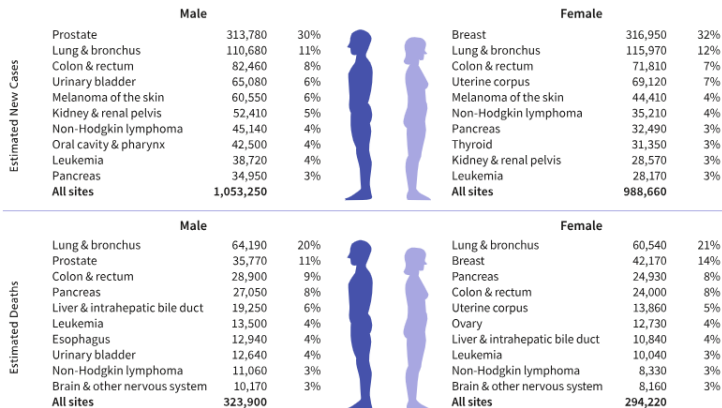


Figure: 2025 年美國癌症統計資料

Siegel, Rebecca L et al. "Cancer Statistics, 2025." CA : a cancer journal for clinicians. 75.1 (2025): 10–45. Web.

癌症的分類

- 小細胞肺癌 (Small-cell lung carcinoma, SCLC) :
 - 佔肺癌的約 15 %。
 - 通常與吸煙有關。
 - 生長迅速，易於轉移。
- 非小細胞肺癌 (Non-small-cell lung carcinoma, NSCLC) :
 - 佔肺癌的約 85 %。
 - 包括腺癌、鱗狀細胞癌和大細胞癌等類型。
 - 生長較慢，預後較好。

癌症的治療方法

- 手術：切除腫瘤組織。
- 放射治療：使用高能輻射殺死癌細胞。
- 化學治療：使用藥物殺死癌細胞。
- 免疫治療：利用免疫系統對抗癌症。
- 靶向治療：針對特定分子或基因突變進行治療。

免疫療法的限制

免疫治療的原理是隔斷 PD-1 與 PD-L1 的結合，解除免疫抑制，使 T 細胞重新啟動並攻擊腫瘤細胞，因此細胞表面的 PD-L1 會直接影響治療的結果，判斷 PD-L1 表現的準確性對於免疫療法成功至關重要。小細胞肺癌患者的 PD-L1 表現通常較低，且對免疫治療的反應較差，因此在臨床上，非小細胞肺癌患者的 PD-L1 表現預測更為重要。

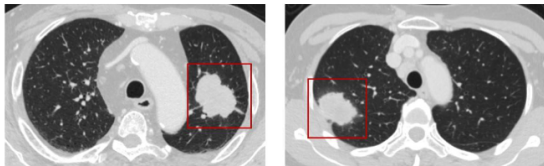


Figure: 左：PD-L1 表現 >50%；右：PD-L1 表現 <50%

圖片來源：周姍妤學姐碩士論文

Section 2

研究目標

研究目標

- ❶ 建立以 MTMAE 為基礎之 PD-L1 表現預測模型
- ❷ 探討加入對比學習對自監督表徵學習的增強效果。
- ❸ 在 ViT encoder 中嵌入 GNN，建構 patch 間關聯性以提升特徵整合能力。
- ❹ 評估多模型集成（ensemble）策略對預測穩定性與泛化能力的影響。

Section 3

研究方法

研究方法

- 實驗材料
- 模型介紹
- 性能指標

實驗材料

- 資料來源：台大醫院、台大醫院新竹分院、台大醫院雲林分院
- 資料類型：非小細胞肺癌患者 CT 與 PD-L1 標記資料
- 總樣本數：188 例病患
- PD-L1 表現分布：
 - PD-L1 expression $\geq 50\%$ (+)：49 例
 - PD-L1 expression $< 50\%$ (-)：139 例

GAN 生成的影像

為了解決資料稀缺問題，本專案使用實驗室先前所開發之 Gabor-GAN 模型，使用公開資料庫 LIDC-IDRT (Lung Image Database Consortium and Image Database Resource Initiative) 生成 518,064 張 2D 的樣本以進行預訓練。

模型介紹

本專題使用或參考了以下的幾個模型及架構：

- Mask Image Model (MIM)
- Masked Autoencoder (MAE)
- Vision Transformer (ViT)
- Multi-task Masked Autoencoder (MT-MAE)
- Simple Contrastive Learning (SimCLR)
- Global Contrastive Masked Autoencoder (GCMAE)
- Contrastive Masked Autoencoder (CMAE)

Masked Image Model (MIM)

- 分成 pretrain, finetune
- 利用 Transfer Learning 的概念，將 pretrain 的 encoder 應用於下游任務
- pretrain 階段，將輸入影像隨機遮蔽一部分，並預測被遮蔽的部分以學習特徵

Masked Autoencoder (MAE)

- MIM 的一種變體
- 利用 Autoencoder 補全被遮蔽的部分以學習特徵
- 將 pretrain 的 encoder 應用於下游任務（如：應用於 ViT 模型以進行分類任務）

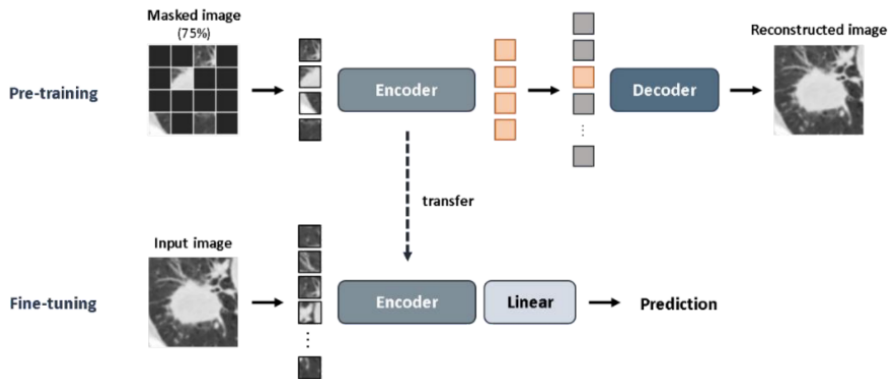


Figure: Masked Autoencoder (MAE) 的架構

圖片來源：周姍妤學姐的碩士論文

Vision Transformer (ViT)

- 將影像分割成 patches，並將其視為序列輸入到 Transformer 模型中
- 利用自注意力機制學習影像特徵
- 使用 CLS token 或是 GAP 處理 token 之後丟到 linear layer 進行分類

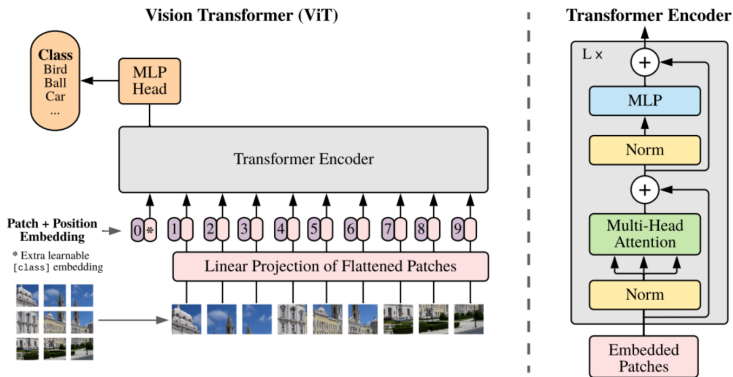


Figure: Vision Transformer (ViT) 的架構

圖片來源：<https://arxiv.org/abs/2010.11929>

Multi-task Masked Autoencoder (MT-MAE)

- 使用大量 GAN 生成的影像進行 pretrain
- 在 pretrain 階段將 MAE 與分割任務結合，使用混合的 Loss 進行優化
- 使用訓練好的 encoder 作為下游分割任務的 backbone

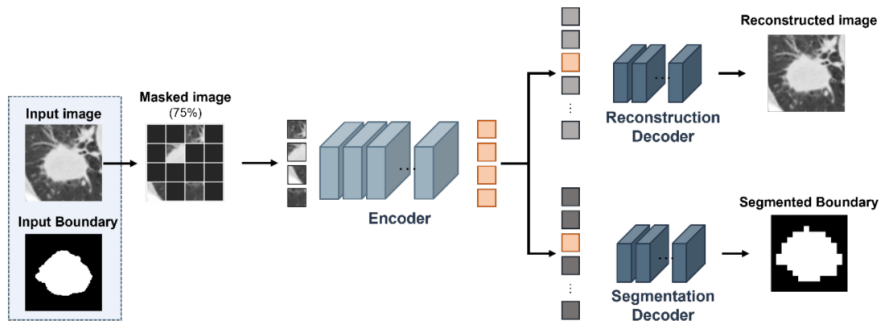


Figure: Multi-task Masked Autoencoder (MT-MAE) 的架構

圖片來源：周姍好學姐的碩士論文

Simple Contrastive Learning (SimCLR)

- 透過對比學習學習影像特徵
- 對同一張影像進行不同的增強，並將其視為正樣本；將不同的影像視為負樣本，使模型學習拉近正樣本，遠離負樣本
- 使用 NT-Xent loss 進行優化

$$\ell_{i,j} = -\log \frac{\exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_j) / \tau)}{\sum_{k=1}^{2N} \mathbb{1}_{k \neq i} \exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_k) / \tau)}$$

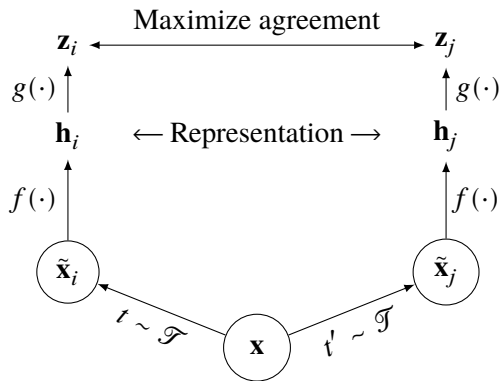


Figure: SimCLR 架構圖

Global Contrastive Masked Autoencoder (GCMAE)

- 結合 MAE 與 SimCLR 的思想
- 利用 MAE 學習局部特徵；結合 GCLR 學習全局特徵

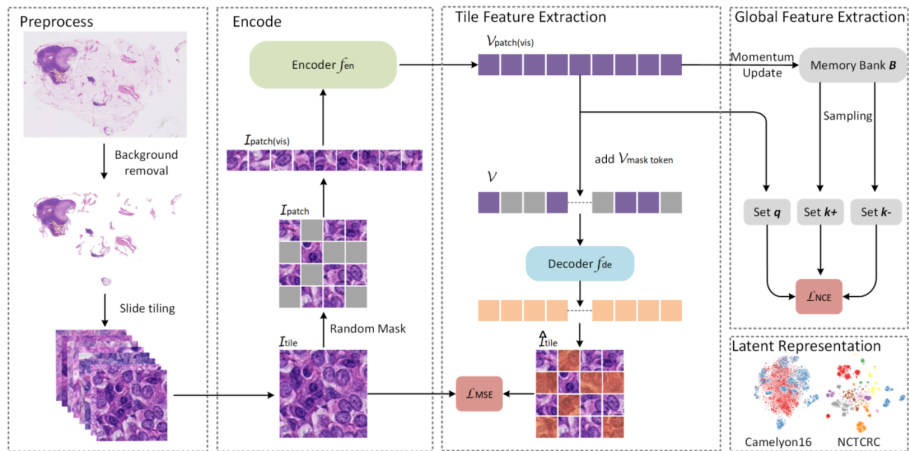


Figure: Global Contrastive Masked Autoencoder (GCMAE) 的架構

圖片來源：<https://arxiv.org/abs/2205.09048>

Contrastive Masked Autoencoder (CMAE)

- 一樣結合 MAE 與對比學習
- 利用孿生網路結構，結合 MAE 與 CL

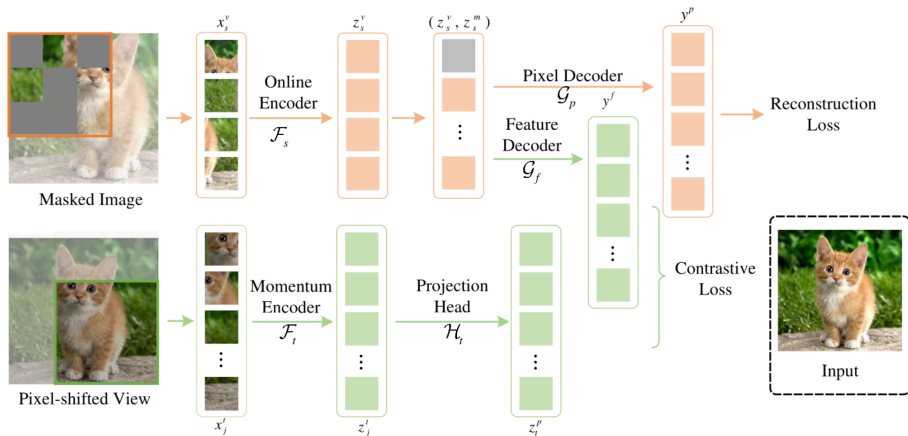


Figure: Contrastive Masked Autoencoder (CMAE) 的架構

圖片來源：<https://arxiv.org/abs/2207.13532>

性能指標

- 正確率 (Accuracy)：預測正確的樣本數與總樣本數之比。
- 靈敏度 (Sensitivity)：預測為正樣本的實際正樣本數與總實際正樣本數之比。
- 特異度 (Specificity)：預測為負樣本的實際負樣本數與總實際負樣本數之比。
- AUC (Area Under the Curve)：ROC 曲線下的面積，用於評估模型在不同閾值下的分類性能。

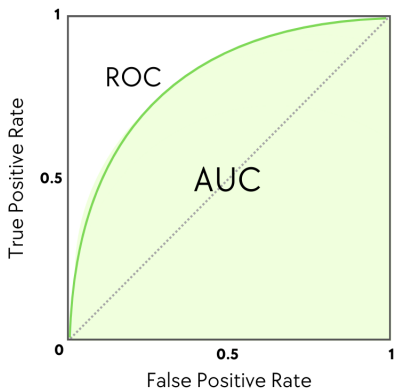


Figure: ROC 曲線及其 AUC 值

圖片來源：<https://www.blog.trainindata.com/auc-roc-analysis/>

Section 4

研究結果與討論

對照組-還原論文的結果

- 參考周姍妤學姐的碩士論文，重現 MTMAE 模型的架構與訓練流程：
 - ① 利用 PyTorch 框架實現模型，並使用學姐留下來的 CT 影像與 PD-L1 標記資料進行訓練
 - ② 進行 200 個 epoch 的預訓練之後得到 MTMAE 模型的權重
 - ③ 使用真實的醫學影像進行微調，並評估模型在 PD-L1 表現預測任務上的準確率
- 最終多次實驗得到的 AUC 平均值為 0.6168
- 結果與學姐論文中所報告相差甚遠，推測是因為資料量太少，不同的排列順序會導致模型預測結果有很大差異

微調的方法

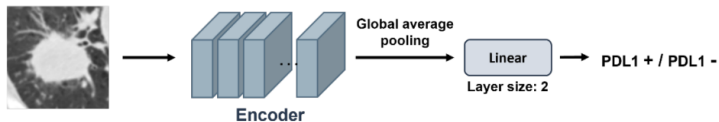


Figure: 微調階段的架構

SimCLR 的訓練



Figure: SimCLR 的訓練方式

SimCLR 的訓練結果

	純粹使用 simCLR	MTMAE + SimCLR 微調
AUC	0.4395	0.5000
ACC	0.4839	0.4355

Table: SimCLR 方法實驗結果比較

由結果可知，純粹使用 SimCLR 訓練的模型表現較差，可能是破壞了原有的特徵學習結構。查閱資料後，我找到了 CMAE 的方法，這個方法結合了 MAE 與 SimCLR 的優點，能夠更好地學習影像特徵。

CMAE 的訓練

由於完整使用 CMAE 預訓練的時間太長，因此我使用了 CMAE 的預訓練權重，和 MAE 的權重進行對比。這兩個權重都是在 ImageNet 上預訓練的。

	CMAE	MAE
AUC	0.5625	0.5584
ACC	0.7419	0.6452

Table: CMAE 方法實驗結果比較

雖然效果不彰，但仍可看出 CMAE 在某些指標上優於 MAE，或許代表了 CMAE 的潛力。

類似模型集成的方法

	5-fold (輪流遮住資料)	5-fold (訓練資料固定)	150 個 epoch
AUC	0.6984	0.5645	0.6644
ACC	0.7419	0.5421	0.7419

Table: 5-fold 方法實驗結果比較

Section 5

結論與展望

結論與展望

- 結合 SimCLR 與 MAE 的方式有些問題，或許可以參考 CMAE 或是 GCMAE 的方法。
- CMAE 的方法有一定的潛力
- 未來預計嘗試：
 - ① 使用完整的 CMAE 模型進行預訓練
 - ② 使用孿生網路的概念結合 MT-MAE 與 CMAE
 - ③ 讓模型不再僅判斷正負，而是將 PD-L1 的表現分成不同層級
 - ④ 嘗試將正確的 k-fold 模型集成策略應用於微調階段，將多個不同的模型進行集成

參考資料

- ① 周姍姍，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，碩士論文，國立臺灣大學，2024。
- ② Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers). 2019.
- ③ Bao, Hangbo, et al. "Beit: Bert pre-training of image transformers." arXiv preprint arXiv:2106.08254 (2021).
- ④ He, Kaiming, et al. "Masked autoencoders are scalable vision learners." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022.
- ⑤ Chen, Ting, et al. "A simple framework for contrastive learning of visual representations." International conference on machine learning. PmLR, 2020.
- ⑥ Quan, Hao, et al. "Global contrast-masked autoencoders are powerful pathological representation learners." Pattern Recognition 156 (2024): 110745.
- ⑦ Huang, Zhicheng, et al. "Contrastive masked autoencoders are stronger vision learners." IEEE Transactions on Pattern Analysis and Machine Intelligence 46.4 (2023): 2506-2517.