

Final Report: 對於準確性與效率的權衡: TransNeXt, MedViTV2 與 RDNet 於腦瘤 MRI 分類任務之比較研究

團隊成員與學號

M144020016 簡以欣、M144020020 林芷庭、M144020063 林傑暉

1. 專案標題 對於準確性與效率的權衡: TransNeXt, MedViTV2 與 RDNet 於腦瘤 MRI 分類任務之比較研究

1.1 專案摘要

本專案旨在探討深度學習模型在醫療影像分析中「模型準確性」與「運算效率」之間的權衡。目前的相關文獻指出，雖然 CNN 在腦瘤 MRI 分類上表現卓越，但在資源有限的臨床環境或行動設備上，運算效率同樣是關鍵考量。

我們選定了三種具代表性的新一代模型架構進行實作與比較：

1. **TransNeXt**: 模擬人類視覺的聚合注意力機制，解決深度退化問題。
2. **MedViTV2**: 專為醫學影像設計，結合 CNN 與 Transformer 的混合架構。
3. **Revitalized DenseNet (RDNet)**: 透過現代化設計重振 DenseNet 潛力的高效模型。

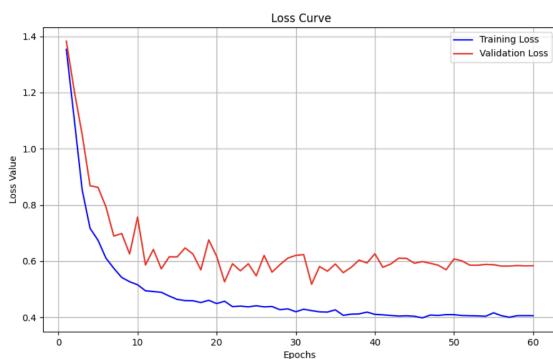
1.2 發現

我們的主要目標是在同一個腦瘤 MRI 任務上，量化比較不同規模模型在準確度表現以及運算效率上的差異。實驗結果顯示，使用參數量較大的模型，在資料集有限的情況下並未能顯著提升模型的辨識能力，反而因為參數量過大導致訓練過程不穩定。

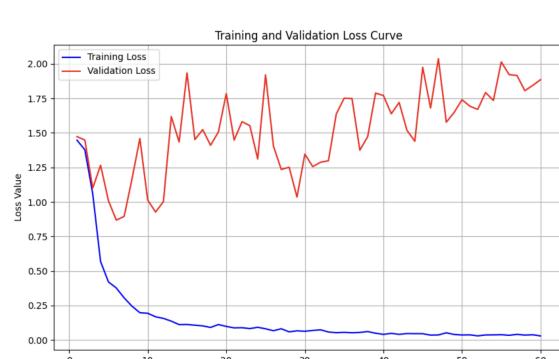
如圖所示，在相同的超參數下，RDNet Tiny 模型雖然學習速度略慢，但是訓練及和驗證集 loss 差距始終在較低的範圍，顯示其泛化能力高，相反的，RDNet Base 模型雖然在 Training Loss 上的下降速度極快，甚至迅速趨近於零，顯示其具備強大的特徵記憶能力；然而，其驗證集 Validation Loss 却呈現劇烈的震盪且不降反升。這種訓練與驗證曲線之間的巨大差距，明確顯示模型發生了嚴重的過擬合現象，證實了在資料有限的情況下，過大的參數量反而會記憶雜訊，導致無法有效推論未曾見過的影像。

這證實了「大模型不一定好」的假設。在腦瘤 MRI 這種資料量相對有限的任務中，RDNet Base 因參數量過大而難以收斂至最佳解，反而被資料中的雜訊誤導，導致實際應用效果遠不如輕量級的 RDNet Tiny。

RDNet Tiny



RDNet Base



2. 前言

2.1 問題定義與動機

磁振造影 (MRI) 是診斷腦瘤的黃金標準，然而，傳統的人工判讀流程不僅耗時，更高度依賴放射科醫師的經驗，易受主觀疲勞影響導致診斷一致性降低。儘管深度學習在醫學影像分析上已展現巨大潛力，但在實際的臨床部署中，仍面臨兩大核心挑戰：

1. 硬體資源受限：大多數醫療工作站或邊緣診斷裝置並未配備高階 GPU 叢集，難以負擔傳統 Vision Transformers (ViT) 龐大的參數量與二次方增長的運算複雜度。
2. 數據稀缺與過擬合風險：相較於自然影像大數據，標註完整的醫療影像極為珍貴且稀缺。大型模型在小樣本數據上極易陷入過擬合 (Overfitting)，導致泛化能力低落。

因此，本研究的核心動機在於尋找並驗證一種高效能輕量化架構。該模型需在維持高診斷準確度的同時，具備極低的運算延遲與優異的數據使用效率，以適應資源受限的真實醫療場景。

2.2 背景與相關工作

傳統 CNN 是醫學影像的主流，但近期隨著 Vision Transformers 的興起，學界開始探索如何結合 CNN 的效率與 ViT 的全局感知能力。為了系統性地評估不同架構設計對腦瘤 MRI 分類的影響，本研究選定以下三種代表 **2024-2025** 年最新技術流派的創新架構進行比較：

2.2.1 TransNeXt **TransNeXt**: 提出了一種模仿生物視覺機制的解決方案，旨在解決深層網路中的特徵退化問題。

- 核心機制：引入「聚合注意力 (Aggregated Attention)」，模擬人類中央凹視覺，透過雙路徑機制在單一操作中同時實現局部細節聚焦與全域資訊捕捉。
- 效率優勢：搭配新型的通道混合器 Convolutional GLU，該模型在降低運算複雜度的同時，顯著提升了對腦瘤邊界等細微特徵的魯棒性。

MedViTV2: 醫學專用混合架構不同於通用的視覺模型，MedViTV2 是專為醫學影像特性設計的階層式混合架構，旨在解決高解析度醫療影像中的特徵崩潰問題。

- 核心機制：它是首批將 Kolmogorov-Arnold Network (KAN) 層引入 Transformer 前饋網路路徑的模型，相比傳統 MLP 具有更高的參數效率與可解釋性。同時，利用 Dilated Neighborhood Attention 來捕捉全域上下文。
- 效率優勢：透過階層式混合策略，MedViTV2 結合了 CNN 的局部特徵提取與 Transformer 的全局感知能力，旨在平衡本地細節與全局關聯性的提取。

Revitalized DenseNet (RDNet) : 在 Transformer 主導的時代，RDNet 重新挖掘了密集連接卷積網路的潛力，挑戰了 ResNet 風格架構的主流地位。

- 核心機制：透過實證研究揭示，特徵的串聯在特徵重用與表達能力上優於加法捷徑。RDNet 透過模塊重設計與寬度擴展，解決了傳統 DenseNet 訓練困難的問題。
- 效率優勢：RDNet 採用了高記憶體效率的現代化設計，顯著降低了訓練與推論時的 GPU 記憶體峰值佔用，使其成為顯存受限環境下的強力候選者。

2.3 創新點

- 首度應用與性能驗證：本研究率先將 TransNeXt、MedViTV2 與 RDNet 三種代表性的新一代圖像分類架構應用於腦瘤 MRI 分類任務。透過在小規模、真實世界醫療數據集上的實證，驗證了這些模型相較於傳統基準模型，在準確度、特徵表達能力與泛化性能上的優越性。
- 效率-準確度系統性權衡分析：針對臨床邊緣運算的需求，本研究建立了一套包含訓練收斂速度、推論延遲與準確度上限的多維度評估基準。我們不單純追求最高的準確度，而是致力於識別出在單位運算成本下效益最高的模型，為醫療 AI 的模型選型提供量化依據。

- 針對醫療影像特性的優化策略：不同於傳統自然圖像分類任務，本研究提出了一套針對 MRI 特性的訓練優化流程。包括採用先篩選後擴展的模型搜索策略，以及針對腦部解剖結構制定的非破壞性資料增強方案，從而成功突破了小樣本數據的過擬合瓶頸。

3. 實驗方法

3.1 資料集與預處理

- 資料來源與分佈：本研究採用公開的 **Kaggle Brain Tumor MRI Dataset**，該資料集包含四種腦部狀態：**Glioma**（膠質瘤）、**Meningioma**（腦膜瘤）、**Pituitary**（腦下垂體瘤）以及 **No Tumor**（無腫瘤）。
- 影像預處理：所有輸入影像統一調整大小至 224×224 像素以符合大多數預訓練模型的輸入規格。在 RDNet-Base 的準確度上限探索階段，我們進一步將解析度提升至 384×384 ，以保留更細微的紋理特徵。標準化 (Normalization) 採用 ImageNet 的平均值與標準差。
- 領域專用數據增強策略 (**Domain-Specific Augmentation**)：由於醫療標註數據稀缺，為防止過擬合，我們實施了針對 MRI 特性優化的增強策略：
 - 幾何與光度增強：啟用 **RandAugment (Magnitude=9)**。此策略引入隨機旋轉、剪切變換 (Shear) 與對比度調整。這模擬了臨床情境中病人頭部擺位差異與不同 MRI 機型的成像參數變化，有效提升了模型對幾何位置與訊號強度的不變性。
 - 停用破壞性混合：刻意將 **Mixup** 與 **CutMix** 係數設為 0。雖然這類技術在自然圖像任務中能提升性能，但在醫學影像中，Mixup 產生的「像素疊加鬼影」以及 CutMix 造成的「解剖結構不連續性」，極易破壞腫瘤的邊界定義。為了確保模型學習到的是真實且具備病理意義的特徵，我們排除了這類可能引入非自然偽影的增強手段。

3.2 系統架構與實作

硬體環境演進：本研究所有實驗均於 Linux 實驗室伺服器環境下進行。考量到 TransNeXt 與 MedViTV2 等新型架構對顯存的高需求，以及為了確保 Batch Norm 層統計估計的穩定性，本專案將硬體配置從初始的單卡 GTX 1080 升級為 2080 Ti 雙卡並行配置。

- GPU 加速器： $2 \times$ NVIDIA GeForce RTX 2080 Ti (11GB VRAM per GPU)。
- 並行策略：採用 **DataParallel (DP)** 進行多卡訓練，有效擴大了 Batch Size 並縮短了實驗迭代週期。

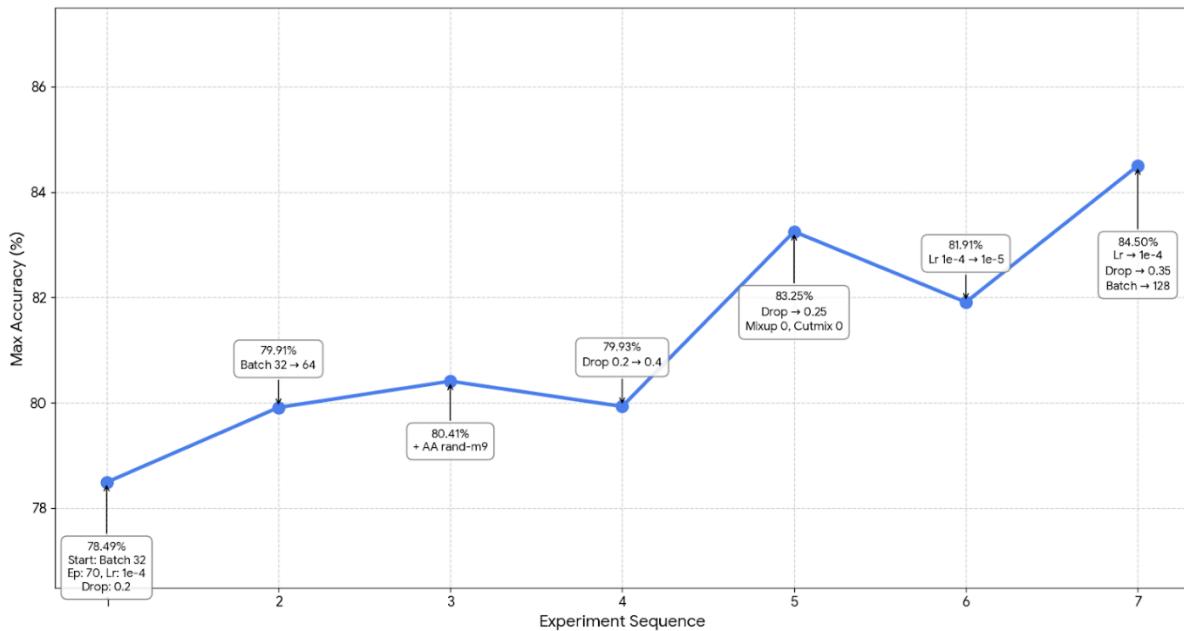
軟體環境：

- 深度學習框架：PyTorch。
- 混合精度訓練：全面啟用 **Automatic Mixed Precision (AMP)**，在保持數值穩定性的前提下，優化記憶體使用率並提升推論速度。

3.3 實作細節

3.3.1 TransNeXt-Tiny

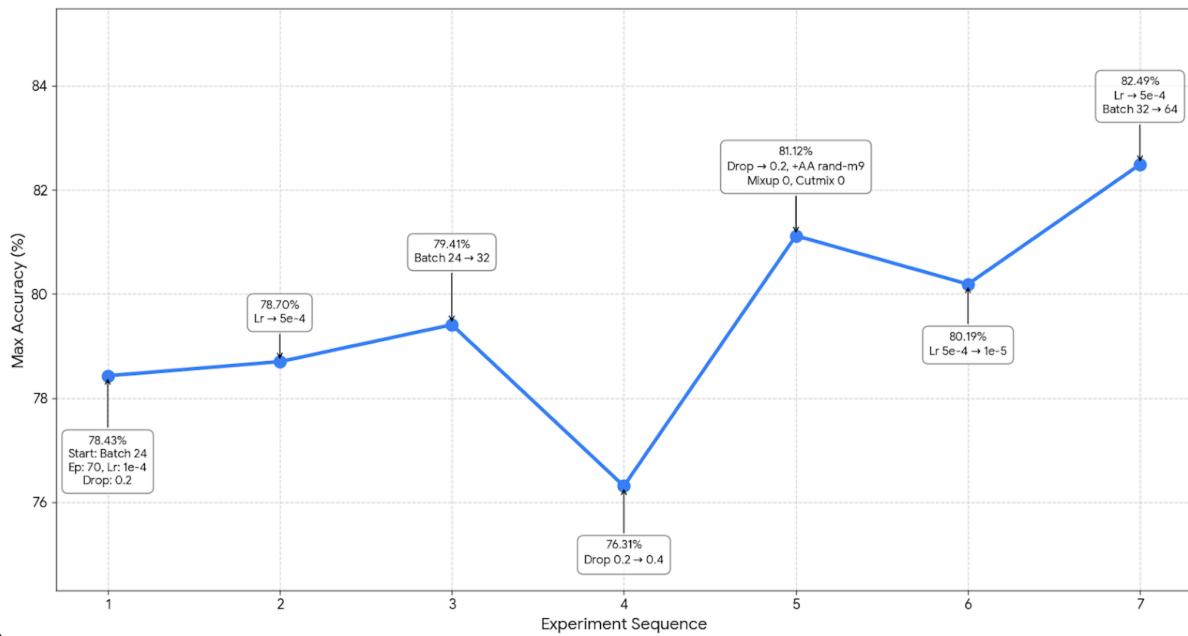
Training Accuracy & Detailed Fine-tuning History



階段	修改內容	準確度	觀察與解釋
初始設定	Batch Size: 32 Epoch: 70 Dropout: 0.2 LR: 1e-4	78.49%	基準模型表現。
修正 1	Batch Size 32->64	79.91%	增加 Batch Size 提升了訓練穩定性或泛化能力，準確度上升約 1.4%。
修正 2	新增 AA rand-m9	80.41%	加入資料增強有效提升了模型強健性，準確度微幅上升。
修正 3	Dropout 0.2 -> 0.4	79.93%	提高 Dropout 導致模型正則化過強，loss曲線略微平穩，但稍微抑制了學習能力，準確度略微下降。
修正 4	Dropout 0.4-> 0.25 Mixup -> 0.0, Cutmix -> 0.0	83.25%	降低 Dropout 並移除 Mixup/Cutmix 後，準確度大幅提升 (+3.3%)。這顯示腦瘤資料集，比較不適合這種破壞性增強。
修正 5	LR 1e-4 -> 1e-5	81.91%	學習率降低，導致模型收斂過慢，準確度回落。
修正 6	LR 1e-5 ->1e-4 Dropout 0.25 -> 0.35 Batch Size 64 -> 128	84.50%	將學習率調回原先較有效率的數值，同時進一步增大 Batch Size 至 128 並適度增加 Dropout。 最終讓訓練集與驗證集 loss 差異約 0.2 且準確性顯著上升

3.3.2 MedViTV2-Small

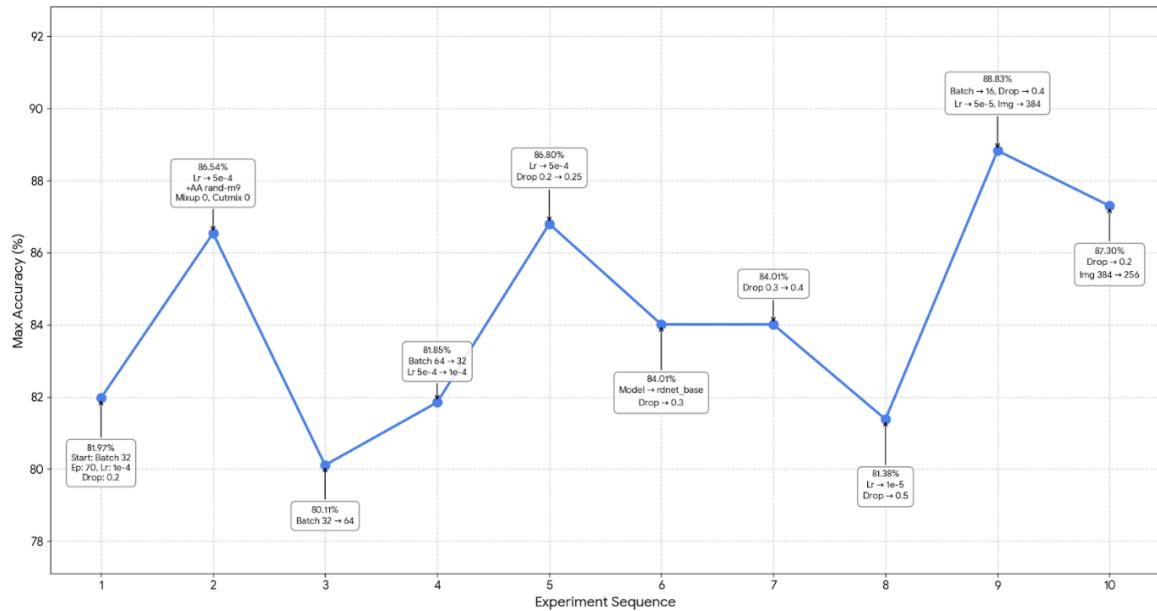
Training Accuracy & Detailed Fine-tuning History (Set 2)



階段	修改內容	準確度	觀察與解釋
初始設定	Batch Size: 24 Epoch: 70 Drop: 0.2 LR: 1e-4	78.43%	基準模型表現。
修正1	LR: 5e-5	78.70%	稍微調低學習率帶來了微幅提升。
修正2	Batch Size 24->32	79.41%	加大 Batch Size 使訓練過程更穩定，從而提升了準確度。
修正3	Dropout 0.2 -> 0.4	76.31%	提高 Dropout 導致模型正則化過強，導致無法學習到足夠的特徵，準確度大幅下降。
修正4	Dropout 0.4-> 0.2 新增 AA rand-m9 Mixup -> 0.0, Cutmix -> 0.0	81.12%	新增資料增強，且避免過度破壞，大幅提升了準確性。
修正5	LR 5e-5 -> 1e-5	80.19%	再次調低學習率，但準確度回落。
修正6	LR 1e-5 -> 5e-5 Batch Size 32->64	82.49%	回調學習率調回原先較有效率的數值，同時增大 Batch Size，準確性顯著上升

3.3.3 RDNet-Tiny、Base

Training Accuracy & Detailed Fine-tuning History (Set 3)



階段	修改內容	準確度	觀察與解釋
初始設定	Batch Size: 32 Epoch: 70 Dropout: 0.2 LR: 1e-4	81.97%	基準模型表現。
修正1	LR 1e-4 -> 5e-5 新增 AA rand-m9 Mixup -> 0.0, Cutmix -> 0.0	86.54%	調低學習率使模型收斂更平緩，同時增加資料增強，準確性大幅上升。
修正2	Batch Size 32->64	80.11 %	加大Batch Size疑似造成模型泛化能力下降，準確性大幅減少。
修正3	Batch Size 64->32 LR 5e-5 -> 1e-5	81.85%	回調Batch Size調低學習率，模型學習緩慢準確度略低。
修正4	LR 1e-5 -> 5e-5 Dropout 0.2->0.25	86.80%	回調學習率，些微提高Dropout，loss曲線更加平緩準確性達到峰值。
修正5	model rdnet_base Drop 0.25->0.3	84.01%	開始測試Base模型，相同參數下效果卻不如Tiny模型，且loss差去大
修正6	Dropout 0.3->0.4	84.01%	提高Drop，減少overfitting

修正 7	LR 5e-4->1e-4 Dropout 0.4->0.5		在次提高正則化，模型曲線趨於平緩但準確性下降。
修正 8	batch-size 32 -> 16 Dropout 0.5->0.4 LR 1e-5 -> 5e-5 img-size 224->384	88.83	調大img-size，使模型看清更多細節，最大準確性上升，但模型極度陡峭。
修正 9	img-size 384->256	87.30	調降img-size，效果差不多

4. 評估方式

鑑於醫學影像數據集(MRI)普遍存在樣本稀缺的特性，直接部署高容量的大型模型極易導致過擬合並陷入局部最優解。此外，考量到臨床邊緣裝置的硬體限制，盲目追求參數量並非最佳策略。因此，本研究制定了由下而上的實驗路徑：優先在 TransNeXt, MedViTV2 與 RDNet 三種架構的Tiny/Small輕量級版本上進行深度優化與歸納偏差驗證，待確立最佳架構後，再進行模型規模的擴展。

4.1 綜合效益評估策略

本階段旨在量化分析三種不同歸納偏差在資源受限環境下的表現。我們引入多維度指標來建構效率-準確度的評估基準：

4.1.1 準確度指標

- **Top-1 Accuracy**: 作為衡量模型分類正確性的核心指標。
- **Stability**: 觀察 Validation Loss 的震盪幅度，評估模型對特徵學習的魯棒性。

4.1.2 效率指標

- **Total Training Time**: 在相同環境下的完整訓練週期時長，反映計算複雜度。
- **Time-to-Convergence** (收斂速率): 模型達到次優解所需的 Epoch 數，這對於需要頻繁重訓練的醫療場景至關重要。

4.2 準確性上限探索

在確認了各架構的效率特性後，第二階段將暫時排除運算成本的硬約束，專注於挖掘特定架構在該數據集上的性能上限。

- **架構篩選**: 基於 4.1 的評估結果，篩選出在輕量級版本中準確度最高且訓練動態最穩定的架構。我們假設：若某種架構的歸納偏差在小參數下能有效捕捉腦瘤特徵，則其擴展版本將具有最高的潛力。
- **模型擴展**: 將該最佳架構擴展至參數量較大的Base或Large變體，以獲得更強的特徵提取能力。

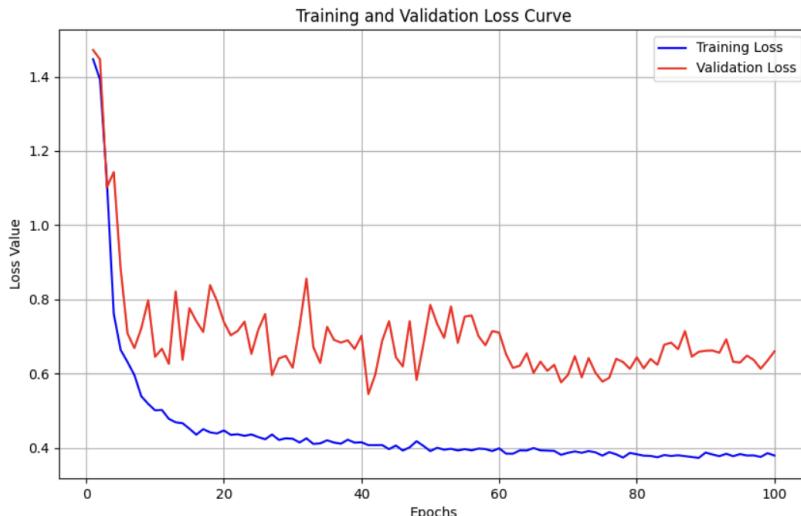
5. 實驗

5.1 綜合性能測試

為了深入理解三種架構在腦瘤 MRI 數據集上的學習行為，我們詳細監控了各模型在訓練過程中的 Loss 下降曲線、驗證集準確度以及收斂速度。以下針對 TransNeXt、MedViTV2 與 RDNet 的訓練結果進行深入解析：

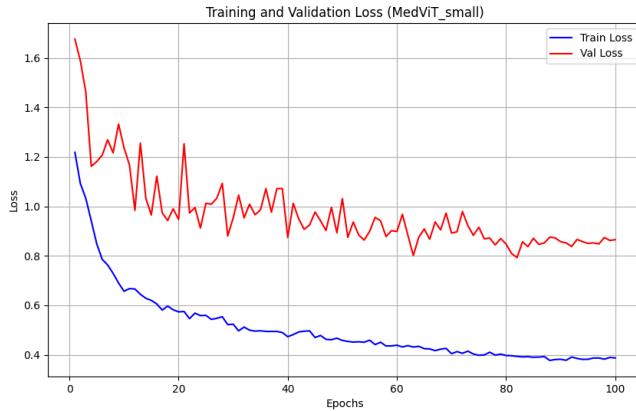
5.1.1 TransNeXt-Tiny: 效率與準確度的最佳平衡

- 實驗數據：Max Accuracy **84.50%** | Training Time **69 mins** (次高) | Epochs 100
- 曲線特徵分析：根據 Loss 曲線，TransNeXt 展現了極為陡峭的初始下降趨勢 (Steep Initial Descent)，Training Loss 在前 5 個 Epoch 內迅速收斂。然而，Validation Loss 呈現顯著的震盪與波動 (**High Oscillation**)。
- 模型行為解讀：
 - 快速探索：得益於其聚合注意力機制，模型能迅速捕捉圖像中的全局特徵。
 - 高變異性：Validation Loss 的劇烈波動顯示模型在圖像特徵中進行激進的探索。雖然穩定性稍遜，但它能在極短時間內達到 **84.50%** 的次高準確度，證明其作為邊緣運算模型具有極高的實用價值，因此我們將其歸類為效率優先場景下的首選。



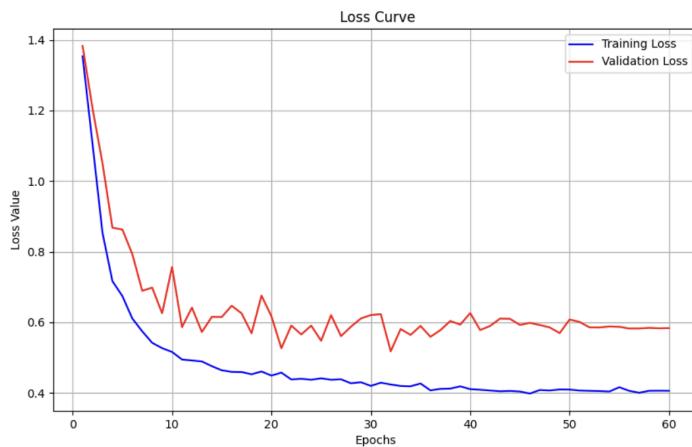
5.1.2 MedViTV2-Small

- 實驗數據：Max Accuracy **82.49%** (最低) | Training Time **74.1 mins** (最慢) | Epochs 100
- 曲線特徵分析：MedViTV2 的表現不如預期。觀察曲線可見顯著的 **Generalization Gap** (泛化差距)。Training Loss 持續下降，但 Validation Loss 始終維持高位且震盪劇烈。
- 模型行為解讀：
 - 特徵學習受限：較大的 Training/Validation Loss 差異暗示模型可能陷入過擬合或難以捕捉具備魯棒性的病灶特徵。
 - 收斂緩慢：混合架構雖然設計精良，但在此特定數據集上顯得較為乏力，需要極高的運算成本卻僅換來最低的準確度。因此，在後續的擴展實驗中將排除此架構。



5.1.3 RDNet-Tiny: 穩定性與準確性之最佳表現

- 實驗數據: Max Accuracy **86.54%** (最高)|Training Time **42.1mins** (最快) | Epochs 60
- 曲線特徵分析: RDNet 展現了最為理想的訓練曲線。Training Loss 呈現平滑且穩定的下降趨勢，且 Validation Loss 在約 **40 Epoch** 後即進入穩定的收斂高原期，沒有出現類似 TransNeXt 的劇烈震盪。
- 模型行為解讀:
 - 特徵復用優勢: RDNet 的密集連接機制有效地促進了梯度的平滑傳播。曲線的穩定性表明模型成功學習到了最具判別力且穩健的特徵表示。
 - 潛力驗證: 在僅 60 個 Epoch 內即達到 **86.54%** 的最高準確度，且曲線末端仍未見明顯過擬合跡象。這強烈暗示該架構在增加參數量(Scaling Up)後，具備最高的準確度上限潛力。



5.2 準確度上限測試

5.2.1 初期瓶頸與挑戰

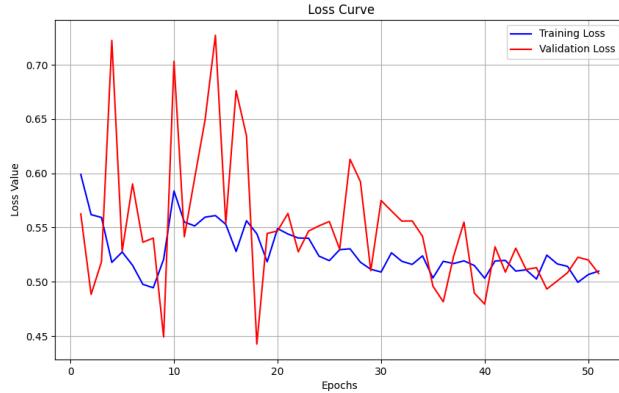
在確認 RDNet-Tiny 為最具潛力的輕量級架構後，本階段實驗旨在透過擴大模型規模至 **RDNet-Base**，以探索該架構在腦瘤 MRI 任務上的準確度天花板。然而，初期將模型從 Tiny 擴展至 Base 時，並未如預期帶來性能提升。受限於數據集規模，大模型出現了過度擬合 跡象；即便加強正規化參數，其準確度在多次實驗中仍低於 Tiny 版本的 86.54%。

5.2.2 優化策略的調整與突破

優化策略的調整與突破 針對上述瓶頸，我們重新審視了訓練策略，推測問題核心並非單純源於模型過大，而是優化超參數與模型容量的不匹配。因此，本研究實施了以下關鍵調整

- 高解析度輸入：將影像尺寸從 224 提升至 384，提供更豐富的紋理細節。
- 微調學習率：大幅降低學習率並配合 Cosine Scheduler，以適應大模型更險峻的 Loss 地貌。

經過此番調整，RDNet-Base 成功突破瓶頸，達到了全專案最高的 88.58% 準確度。



6. 討論與結論

6.1 綜合性能評估

在量化數據中，RDNet-Tiny 憑藉 42 分鐘的總訓練時間看似在效率上略勝一籌。然而，若深入分析訓練動態，TransNeXt-Tiny 在特定應用場景下展現了更優越的探索效率

- 收斂速度：TransNeXt 的 Loss 曲線在前 5 個 Epoch 呈現垂直式下降，並在 **30-40 Epochs** 左右即達到高原期。這意味著該模型具備極強的特徵搜尋能力。
- 早期停止的適用性：RDNet 的 Loss 曲線相對平緩，顯示其依賴層層堆疊的密集卷積來逐步完善特徵，若過早結束訓練，模型可能尚未建立穩健的特徵表達。相反，TransNeXt 若採用早期停止策略，僅需約 **30%** 的總訓練預算即可達到可用的準確度水平。
- 結論：若以完成完整訓練為標準，RDNet 最優，但若以在最短時間內獲得可用模型為標準，TransNeXt-Tiny 提供了最佳的效率與準確度平衡。

6.2 準確度上限與可靠性評估

6.2.1 特徵飽和假設的再檢視

前期實驗中，即使經過多輪優化，Base 模型常難以顯著超越 Tiny 模型。這支持了我們最初的假設：腦瘤 MRI 數據集的特徵複雜度有限，Tiny 級別的模型容量 (Model Capacity) 已足以提取所有顯著特徵，更大的參數空間反而引入了優化雜訊。

6.2.2 優化策略的盲點

後續的驗證中，雖然最終實驗突破了分數限制，但觀察 RDNet-Base 的訓練 Loss 曲線，可發現其呈現顯著的震盪與尖峰，與 Tiny 版本的平滑曲線形成強烈對比，對此我們對於，該結果抱持強烈的存疑，並且提出造成此現象的可能原因。

- 去耦合現象：觀察訓練過程，Training Loss 與 Validation Loss 呈現極不規則的高頻震盪。更關鍵的是，兩者出現了明顯的去耦合現象：當 Validation Loss 落入最低點時，Training Loss 並未呈現相對應的顯著下降或收斂跡象，且兩者間的泛化差距極大。這顯示模型在該時間點的優異表現，並非源於對訓練數據特徵的穩健掌握。
- 隨機對齊：我們推測該模型並未真正收斂至一個具備魯棒性的全域極小值，相反地，模型似乎處於一種隨機猜測的狀態，Loss 曲線在低谷後的迅速反彈，證實了該最優解極不穩定。這暗示模型權重只是在激烈的參數更新中，剛好在某個瞬間與驗證集的特徵分佈產生了巧合性的對齊。

6.2.3 最終結論

綜上所述，儘管 RDNet-Base 取得了 88.58% 的數據突破，但其背後的學習機制顯示出脆弱性與投機性，難以保證臨床應用時的泛化能力。

因此，本研究最終結論為：RDNet-Tiny (86.54%) 雖然準確度略低 2%，但憑藉其優異的訓練穩定性、紮實的特徵學習曲線以及較低的運算需求，相比於不穩定的 Base 版本，才是更具備真實臨床部署價值的最佳選擇。這也驗證了在醫療 AI 領域，模型的可信度與穩定性往往優於單純的準確度峰值。

7. 個人貢獻聲明

- 簡以欣 (36%): 負責 RDNet 模型的實作、Tiny 與 Base 版本的超參數調整、Mixup/Cutmix 進階增強策略的實現與分析、最終報告 ppt、文稿。
- 林芷庭 (28%): 負責 MedViTV2 模型的環境架設與超參數調整。
- 林傑暉 (36%): 負責 TransNeXt 模型的移植、解決 Loss 震盪問題、數據集預處理、最終報告分析、整合及 pdf。

8. 參考文獻

- [1] Kaggle Dataset: Brain Tumor Classification (MRI).
- [2] Shi, D. (2024). TransNeXt: Robust Foveal Visual Perception for Vision Transformers. arXiv preprint arXiv:2311.17132.
- [3] Manzari, O. N., et al. (2025). Medical Image Classification with KAN-Integrated Transformers. arXiv preprint arXiv:2502.13693.
- [4] Kim, D., Heo, B., & Han, D. (2024). DenseNets Reloaded: Paradigm Shift Beyond ResNets and ViTs. arXiv preprint arXiv:2403.19588.