元智大學資訊管理系

學術類畢業專題頂石課程(二)

期末報告

跨越偏鄉的診斷革命:
生成式 AI 在白血球分類中的應用

1111624 游凱甯、1111735 李秉霖
1111729 廖奕玟、1111742 陳薇珊、1111723 江昀芮

實習公司：ZI1-謝瑞建
指導教授：謝瑞建 博士

中華民國 114 年 11 月
Nov, 2025

Abstract

本提案針對白血球影像的分類問題，利用深度學習技術實現 14 種白血球類別的精準識別。研究資料來自衛生福利部桃園醫院檢驗科，涵蓋 14 種類別臨床白血球影像。由於部分類型白血球細胞資料樣本不足，模型無法進行訓練，因此我們使用 Diffusion-GAN 生成之模擬資料，且生成影像的真實性與多樣性採用多項指標以及獲得國家認證之細胞型態分類臨床專家人員評估認證。此外，我們透過 APP 將血液細胞影像拍攝並上傳至本提案架構的 Clip-LLaMA 伺服器，利用影像及語意的配對進行，產生白血球的分類結果。此架構與過往透過影像特徵進行分類大不相同，充分利用更為創新的 Text as Hub 架構技術，透過語意進行模型分類，且在分類白血球細胞上的準確率達 89%，超越我們邀請的三位獲得國家認證之細胞型態分類臨床專家人員的辨識能力，能有效解決各醫院獲得國家認證之細胞型態分類臨床專家人員不足的問題，有效減緩資深檢驗人員的工作負擔。此外，結合的大型語言模型能夠輸出分類依據，輸出的分類依據經由多項指標評估達標，讓資淺的檢驗人員在分類的同時能自我學習。研究成果已驗證分類的準確性能分擔資深檢驗人員的工作，輸出分類依據也能夠達到教育訓練的目的。

關鍵詞：CLIP, LLaMA, Text as Hub, 白血球影像分類

Contents

Chapter 1 前言

白血病(Leukemia)，俗稱血癌，對於其預防與治療，白血球的分類扮演著至關重要的角色。白血球在血液中的種類與分佈變化，能夠提供醫生診斷疾病的重要線索，尤其是在血癌的早期篩查與診斷過程中，準確的白血球分類至關重要。根據台灣癌症基金會的資料，台灣的十大死因中，癌症高居首位，其中淋巴癌(惡性淋巴瘤)及白血病分別位列癌症死亡率的第八和第九位，尤其是小兒血癌更為小兒癌病的首位，顯示出血癌對於台灣人民健康的嚴重威脅，因此，準確辨識白血球細胞，將是提升疾病預防與治療成效的關鍵。

然而，臨床上使用的血液檢測設備往往價格昂貴，僅有大城市的醫院能夠負擔得起，對於規模較小或偏遠的醫療機構來說，這些先進設備難以普及，因此他們通常依賴人工觀察血液玻片來進行診斷。在這些醫療機構中，由於部分檢驗人員的年資較短，接觸的病例數量不足，導致臨床經驗的相對匱乏。此時，檢驗人員需要依賴與總院之間的聯繫協助，以便進行更為準確的診斷。然而這一過程不僅耗時也極易出錯，從而產生不容忽視的隱形成本。因此如何在偏遠地區的醫院中提高診斷效率與準確性，成為當前必須解決的問題。

本提案所取資料集的 14 類白血球細胞，為臨床認列檢驗人員例行事務需要分辨的 14 種類別。本提案採用衛生福利部桃園醫院醫事檢驗科所提供的資料，但因某些類別較為罕見，造成資料量樣本不足，模型無法進行訓練，於是本提案引入Diffusion-GAN技術來生成白血球影像，生成的影像經獲得國家認證之細胞型態分類臨床專家人員認證，與真實影像相差無幾，成功解決資料量不足的問題。

本提案利用 CLIP 結合 LLaMA 模型，達到分類白血球的目標以外，還能輸出分類依據，讓年資較短的檢驗人員能夠自我學習，達到教育訓練的目的。我們希望能有效減少對昂貴醫療設備與資深檢驗人員的依賴，並為偏遠地區的醫療機構提供更為高效、準確的解決方案。藉由這項提案，我們期望能夠改善當前醫療資源分配不均的現狀，並幫助更多偏遠地區的醫療機構提升診斷水平。

專案架構說明

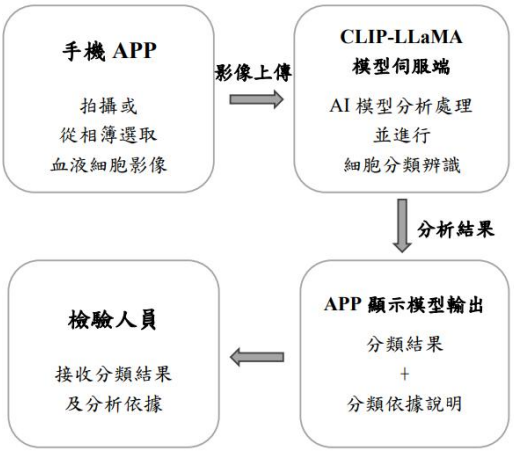


圖 1. 血液細胞影像分析系統架構圖

系統流程說明(如圖 1):

(a) 檢驗人員使用手機 APP 拍攝或選取血液細胞顯微鏡影像

(b) APP 會自動將影像上傳至 CLIP-LLaMA 模型所在的伺服端進行 AI 分析處理

(c) APP 會自動接收模型輸出的細胞分類結果及詳細的分類依據說明並顯示在 APP 上

(d) 檢驗人員查看完整分析結果，協助專業醫療判斷，也可以依據模型輸出的依據達到自我學習，提升分辨細胞的專業能力

Chapetr 2 痛點說明

白血球資料集共涵蓋 14 個類別，其中部分屬於罕見類別，樣本數量較為稀少，不僅使檢驗人員在分類時因經驗不足而面臨困難，也影響深度學習模型的訓練成效。資源有限的中小型醫療機構，由於病患數量有限，造成檢驗人員缺乏足夠的細胞型態分類經驗，需仰賴與總院資深檢驗人員遠距協作，進而增加其工作負擔。此外，血球分類儀器價格昂貴，多數中小型醫院無力購置，即使是大型醫院也多以租賃方式使用，成本相當可觀。

為解決上述問題，本提案使用 Diffusion-GAN 技術進行資料增補，生成資料經 FID、IS、CMMD 等多項評估指標驗證皆具良好表現，並經由三位擁有國家認證之細胞型態分類臨床專家進行判讀測試，顯示其真實性與原始資料難以區分，驗證生成資料的品質具高度可信度。

此外，本提案提出與過往大不相同的 Text as Hub 架構，突破傳統僅依賴影像特徵的分類方式，透過語意進行模型分類。我們透過 APP 將血液細胞影像拍攝並上傳至 CLIP-LLaMA 伺服器，結合影像及語意的配對進行分類，產生白血球的分類結果，準確率超過 89%，優於三位已獲得國家認證的細胞型態分類臨床專家人員的分辨能力，顯示本模型具備實際應用潛力，可分減輕資深檢驗人員之負擔。除此之外，模型所提供的分類依據可作為學習資源，提供臨床人員自我學習的機會，協助檢驗人員補足細胞型態分類經驗，進一步提升臨床診斷效能。

Chapter 3 研究目的

本提案旨在建構一套高效能且具解釋能力的白血球細胞分類系統，協助提升血癌診斷中細胞分類的準確性與應用普及性。資料量不足為白血球影像分類中常見之挑戰之一，部分細胞類別樣本數量偏少，導致模型訓練穩定性與泛化能力受限。為此，本提案運用 Diffusion-GAN 技術生成白血球影像，進行資料增補，以提升訓練資料之多樣性與均衡性，並為後續模型訓練提供堅實基礎。

在資料充分後，本提案以 CLIP ViT-H/14 作為分類模型基礎，結合自訂 MLP 分類器，並透過微調(fine-tuning)方式適應醫學影像特性。經訓練後，本系統於分類任務中展現出優異表現，整體準確率超越通過國家專業認證且具五年以上經驗之資深醫檢人員水準，顯示本方法於臨床輔助判斷之應用潛力。

同時，為提升推論結果之可解釋性與輔助教學功能，本提案引入大型語言模型 LLaMA-3-8B，透過 LoRA 技術進行輕量化微調，建立從影像特徵提取到醫學描述生成之端到端推論流程。結合 Text as Hub 技術進行生成描述之語意品質評估，確保輸出文

本之臨床實用性與準確性，使系統能針對分類結果生成直觀且專業的文字說明，有助於資淺醫檢人員或醫學生之快速理解與學習。

本提案最終目標在於建立一套低成本、高準確度且高可用性的智慧診斷系統，克服傳統細胞分類設備昂貴與人力成本高昂的問題，並可推廣至資源有限之偏遠地區，提升偏鄉醫療機構於血癌早期篩查之診斷效率與整體醫療服務品質。

Chapter 4 文獻探討

4.1 Diffusion-GAN

隨著深度學習技術的進步，醫學影像分析已逐漸由傳統影像處理邁向基於神經網路的智慧辨識，特別在白血球細胞分類領域，多篇研究證實卷積神經網路(CNN)在影像辨識任務中展現高度準確率。如 M. Yıldırım and A. Cinar [16]使用 DenseNet201 模型對 4 種白血球細胞進行分類，達到 83%的準確率，Supawit Vatathanavaro, Suchat Tungjitnob, and Kitsuchart Pasupa [2]則以 ResNet 進行 5 種白血球細胞分類，準確率達 88%。

然而，細胞影像資料通常存在類別分布不均與樣本數不足的問題，進一步影響模型訓練準確度與泛化能力。為解決此問題，近年來有研究採用生成對抗網路(GAN)進行醫學影像資料擴充。Frid-Adar, M., Diamant, I., Klang, E., et al. [6]透過 GAN 生成肝臟病變影像，使分類準確率顯著提升。

本提案結合Diffusion-GAN進行白血球模擬影像生成，補強類別不平衡問題，生成之影像皆經由多項指標驗證，且我們也邀請三位年資五年以上且獲得國家認證之細胞型態分類臨床專家人員進行盲測評估，結果顯示生成影像的真實性極高，人為辨識也難辯其真假。

4.2 CLIP 與 LLaMA 模型

隨著人工智慧與深度學習技術的持續發展，多模態模型(Multimodal Models)結合視覺與語言訊息的能力逐漸受到重視，推動了智慧醫療、智慧教育等領域的創新應用。

Radford 等人提出的 CLIP(Contrastive Language-Image Pretraining)模型[7]，為多模態學習領域的重要里程碑。CLIP 透過大規模圖文配對資料進行自監督對比學習，結合影像編碼器與文字編碼器，並以最大化正確配對相似度、最小化錯誤配對相似度作為訓練目標，使模型能同時理解影像與自然語言。該研究證實，自然語言可作為有效監督信號，大幅提升模型的跨任務遷移能力，並在多項零樣本分類(zero-shot classification)任務中表現出優異的性能。

本提案延續 CLIP 的設計理念，選用社群開源之 CLIP-ViT-H/14 模型作為影像特徵提取器，並針對白血球細胞分類任務進行再微調(fine-tuning)。CLIP-Encoder 採用 Transformer 架構為基礎之設計，該架構由 Vaswani 等人提出[8]，強調以自注意力(Self-Attention)機制捕捉序列中元素間的關聯性，成為現代視覺理解模型的重要基礎。ViT(Vision-Transformer)架構則是由 Dosovitskiy 等人提出[9]，將 Transformer 直接應用於影像領域，突破了傳統卷積網路的侷限，並在多項視覺任務中展現優異表現。同時，為提升模型對序列結構的敏感度，ViT 架構亦融入了 Shaw 等人所提出的相對位置編碼(Relative Position Representations)技術[10]，以加強模型對影像空間結構的感知能力。

此外，CLIP-Adapter 的研究證明，透過在 CLIP 特徵輸出後加接輕量適配器層(Adapter Layers)並進行微調，可有效提升模型在特定任務的表現[11]。本提案亦採用類似概念，在 CLIP Encoder 特徵後接自訂多層感知器(MLP)分類器，並同步微調，以適應醫學影像分類任務之需求。透過此設計，成功提升模型於 14 類白血球細胞分類任務上的準確率，使其在特定領域中展現出媲美資深檢驗人員的判讀能力。

而為了強化推論結果的可解釋性，本提案引入大型語言模型(Large Language Model, LLM)LLaMA-3-8B，並透過 LoRA(Low-Rank Adaptation)技術[12]進行輕量化微調。LoRA 技術能有效減少微調時需要更新的參數量，提升訓練效率並降低硬體需求，適合應用於大型模型之特定任務適配。LLaMA-3-8B 則為 LLaMA 系列之延伸，其系列最初由 Touvron 等人提出 LLaMA v1)[13]，後續經 LLaMA2 改良[14]，並於 2024 年推出 LLaMA3(Llama Team)[15]，在開源社群中展現卓越效能與廣泛應用性。本提案以微調後之 LLaMA-3-8B 模型，將 CLIP 所提取之影像特徵轉化為自然語言描述，並引用 Text as Hub 框架，將影像特徵轉換為白血球細胞醫學描述，使系統除能完成分類任務外，亦能自動生成符合醫學標準之細胞敘述文本，大幅提升推論結果之可讀性與可解釋性。

Chapter 5 研究方法

5.1 資料集介紹

(a) 資料來源：衛生福利部桃園醫院醫事檢驗科並由血液鏡檢組長葉建宏提供。

(b) 數據格式：368*370.jpg 圖像格式。

(c) 標註信息：

(d) 資料特點：涵蓋了來自不同年齡段和性別的患者資料，且不同類別的白血球數量有較大差異。

〔表 1. 資料集各類別數量表〕

細胞種類	**數量**
Smudge Cells	1130
Promyelocyte	1204
NRBC	1756
Neutrophilic Myelocyte	2129
Neutrophilic Metamyelocyte	1849
Neutrophil Segment	11443
Monocyte	11500
Lymphocyte	10394
GIANT PLT	4287

Eosinophils	7427
Blast	5642
Basophils	4652
Band	1626
Atypical lymphocyte	1197

從表 1 可看出，部分類別較為罕見導致樣本數量較少(少於 4000)，在數據不足的情況下，模型的訓練無法進行且分類效果未能達到預期。特別是部分細胞類別之間具有較高的相似性，當這些類別的樣本數量不足時，模型難以有效區分，從而造成預期以外的分類結果。

5.1.1 資料填補

為了解決數據不足導致模型難以訓練，從而造成預期以外的分類結果的問題，我們採用了 Diffusion-GAN 來進行資料填補。

Diffusion-GAN 結合了擴散模型和生成對抗網路的優勢，透過模擬數據的生成過程來創建新的樣本，這種生成方式能夠產生結構和細節上與真實數據相似的樣本，並有效彌補樣本不足的問題。

為確保生成影像品質，我們對所有生成影像進行了多項指標的評估，並請專業檢驗人員對進行主觀評估。（註：*為經過資料填充）

〔表 2. 經資料填充各類別數量表〕

細胞種類	**數量**
Smudge Cells	*5200
Promyelocyte	*5200
NRBC	*5200
Neutrophilic Myelocyte	*5200
Neutrophilic Metamyelocyte	*5200
Neutrophil Segment	11443
Monocyte	11500
Lymphocyte	10394

GIANT PLT	*5200
Eosinophils	7427
Blast	5642
Basophils	*5200
Band	*5200
Atypical lymphocyte	*5200

5.1.2 生成影像之指標評估

本提案使用以下指標來評估生成對抗網路所生成的白血球影像之真實性以及多樣性，並作為衡量生成影像與真實影像之間差距的依據。

(a) FID(Frechet Inception Distance)：

用於衡量生成影像的真實性。FID 是一種衡量生成圖像與真實圖像之間分佈差異的指標。其計算方法依賴於預訓練的 InceptionV3 模型，將圖像轉換為高維特徵向量，並計算這些向量的均值和協方差之間的 Fréchet 距離。FID 的值越低，表示生成圖像與真實圖像在特徵空間中的分佈越接近，從而表現出較高的真實性。

(b) CMMD(Causal Maximum Mean Discrepancy)：

CMMD 是用於衡量兩組圖像分佈差異的指標，根據最大均值差異(MMD)計算生成圖像和真實圖像之間的統計距離。該指標專注於衡量生成圖像與真實圖像在特徵空間中的差異，從而評估生成模型的真實性或多樣性。CMMD 可以有效處理不同類型的生成模型，並在多樣性和真實性之間提供平衡的評估。

5.2 模型

5.2.1 整體模型流程架構

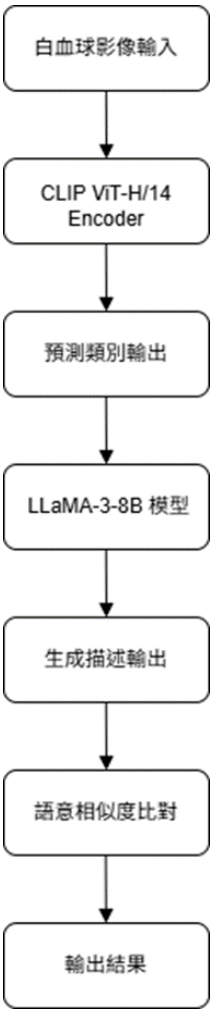


圖 2. 整體模型架構流程圖

將白血球影像輸入微調後的 CLIP ViT-H/14 提取特徵，經由分類器預測白血球類別，並依據預測結果生成對應的提示語句，輸入至微調後的 LLaMA-3-8B 模型，產生該類別的醫學描述。隨後將生成描述與標準描述進行語意相似度比對，最終輸出包含分類結果、描述文字及語意評估的推論成果。

5.2.2 CLIP 模型

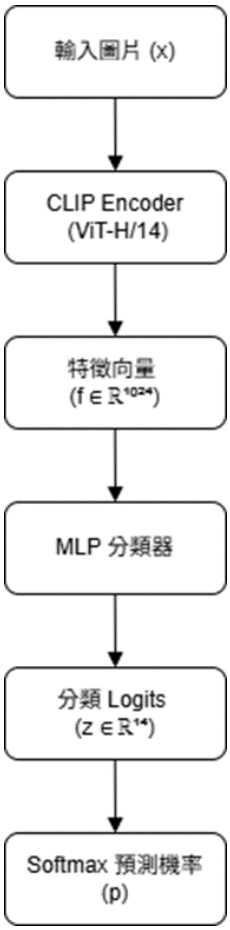


圖 3. 為 CLIP 模型架構圖

本提案以 CLIP ViT-H/14 提取白血球影像特徵，並設計自訂的 MLP 分類器接續於特徵向量之後，透過微調 CLIP Encoder 與 MLP，完成針對白血球的分類任務。

(a) 方法概述

(i) CLIP Image Encoder：提取圖片的深層語意特徵。

(ii) MLP Classifier：基於特徵進行 14 類別的白血球分類。

訓練時，CLIP Encoder 與 MLP 同時進行微調，而非僅調整分類器。

(b) 過程公式推導

(i) 特徵提取

輸入影像$X\epsilon R^{3\times224\times224}$，經由 CLIP=Encoder 抽取特徵向量$f\epsilon R^{d}$，其中 d=1024。

$$f = CLIP - Encoder(\pi) \tag{1}$$

(ii) 分類器輸出

特徵向量 f 經過小型多層感知器(MLP)，輸出分類 logits $z\epsilon R^{14}$:

$$z = MLP(f) \tag{2}$$

（iii）將 logits 經 softmax 得到分類機率，取最大值作為預測類別。

(c) 訓練結果

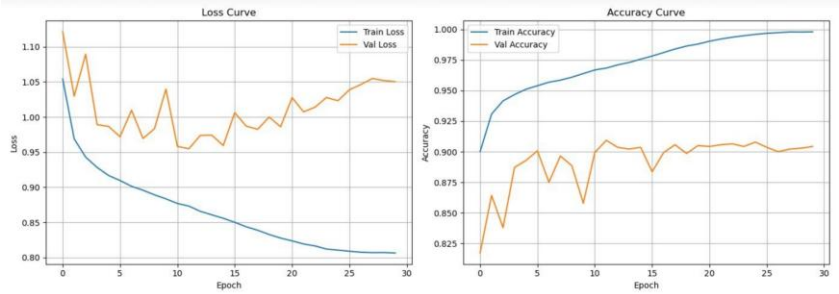


圖 4. 為 CLIP 模型學習曲線

圖為 CLIP 模型於訓練過程中損失值(loss)與準確率(accuracy)隨訓練次數(epoch)變化的曲線。可觀察到模型在訓練期間的收斂情形，包括 loss 及 accuracy 的趨勢，進而評估模型學習過程的穩定性。

5.2.3 LLAMA 模型

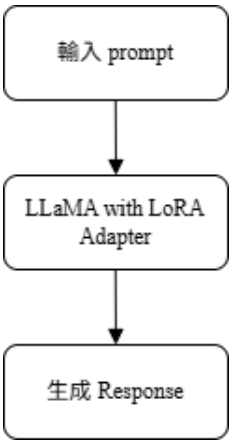


圖 5. 為 LLaMA 模型架構圖

本流程圖說明 LLaMA 模型的設計。首先輸入一組 prompt，透過加掛 LoRA Adapter 的微調後 LLaMA 模型進行處理，最終輸出對應的 Response 作為生成結果。

(a) 目的

影像分類可以辨識白血球的種類，但在實際醫療應用中，生成對應的醫學描述能提供更高層次的解釋性與診斷輔助。因此，針對白血球分類後的類別，我們設計特定的醫學描述，並以輕量化方式微調 LLaMA-3-8B 訓練語言模型，讓模型能根據給定的 prompt，自動生成準確的醫學描述。

(b) 方法概述

（i）基礎模型：LLaMA-3-8B Instruct(4-bit 量化版本)

（ii）微調技術：LoRA(Low-Rank Adaptation)，只調整小量參數

（iii）微調資料：以「prompt ➔ response」格式建立，描述白血球各種類的特徵與疾病意義

(c) 過程公式推導

(i) 基本目標

對於每組(x,y)(prompt 和 reference description)，最小化生成描述$\hat{y}$與標準描述 y 的交叉熵損失：

$$L_{SFT} = -\sum_{t=1}^{T} log\ p(y_t|y < t, x) \tag{3}$$

其中:

x=prompt

y=(y1,y2,...,yT)：正確的 response token 序列

p(・)：模型在每個時間步預測的機率分佈

(ii) LoRA 插入

LoRA 以輕量化方式調整權重，只需少量參數更新，無需調整整個模型，並用「原權重加上低秩權重」概念完成。

(iii) Semantic Accuracy 計算

在驗證時，我們不是用 token-level accuracy，而是使用語意相似度（Semantic Similarity）來評估生成描述與標準答案的一致性。

給定生成文本$\hat{y}$和標準文本 y，用 SentenceTransformer 取 embedding 向量。

相似度計算:

$$Semantic\ Accuracy\ = cos(embedding(\hat{y}) \\ ,embedding(y)) \tag{4}$$

如果平均語意相似度高於門檻值(例如 0.7)，即認定生成品質良好。

(iv) 訓練結果

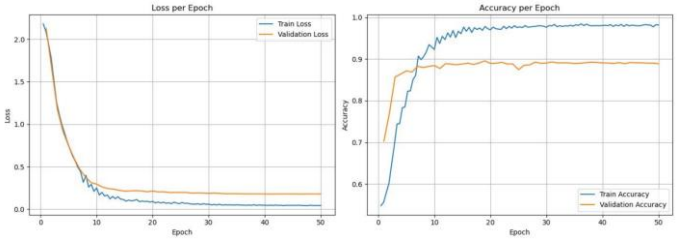


圖 6. 為 CLIP 模型學習曲線

圖為 LLAMA 模型於訓練過程中損失值(loss)與準確率(accuracy)隨訓練輪數(epoch)變化的曲線。圖中可觀察到模型在訓練期間的收斂情形，包括 loss 及 accuracy 的趨勢，進而評估模型學習過程的穩定性。

5.2.4 CLIP 結合 LLAMA 之推論步驟與方式

推論的流程主要包含 3 大步驟:

Step1: 影像分類：利用 Fine-tuned CLIP Encoder + MLP Classifier 進行 14 類分類預測。

Step2: 文字生成：將分類結果填入 Prompt 模板，並輸入 Fine-tuned LLaMA 生成對應描述。

Step3: 相似度評估：使用 Soft Cosine Similarity，計算生成描述與參考描述的語意相似度：

$$SoftCosine(v1, v2) = \frac{v1\mathrm{T}Sv2}{\sqrt{v\mathrm{T}1Sv1}\sqrt{v2\mathrm{T}Sv2}} \tag{5}$$

5.2.5 分類及語言模型之指標評估

本提案使用多項指標來評估白血球影像分類與文字生成模型的整體表現。這些指標分別量化了分類任務中的準確性與錯誤類型，以及生成描述在語意層次的相似度與品質，作為定量分析模型理解與描述白血球影像能力的重要依據。

(a) CLIP + MLP:

(i) Accuracy

(ii) Confusion Matrix

(iii) Precision / Recall / F1-score

(b) LLaMA:

(i) Soft Cosine Similarity

Chapter 6 結果與討論

6.1 生成影像之評估

6.1.1 生成影像之指標評估

本提案透過兩種面向(真實性、多樣性)對共 14 種類別所做出來的生成影像做品質評估。

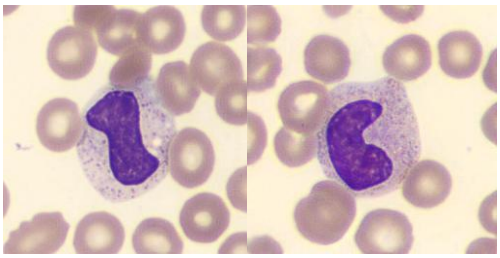


圖 7. 為 Band 之真實(左)以及生成影像(右)對照

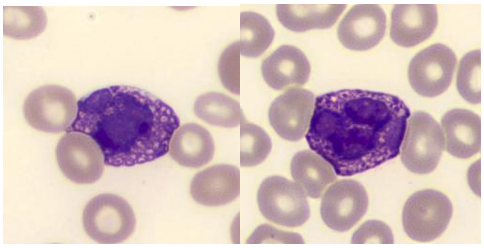


圖 8. 為 Basophils 之真實(左)以及生成影像(右)對照

〔表 3. 生成影像 FID 指標結果〕

細胞種類	**FID**
Smudge Cells	34.33
Promyelocyte	43.13
NRBC	35.88
Neutrophilic Myelocyte	37.52
Neutrophilic Metamyelocyte	35.31
Neutrophil Segment	34.50
Monocyte	32.32
Lymphocyte	31.12
GIANT PLT	38.51
Eosinophils	34.77
Blast	31.44
Basophils	35.89
Band	35.23

Atypical lymphocyte	37.43

根據 FID 指標的結果，Smudge Cells 以 34.33 的 FID 值表現出色，顯示其生成圖像在真實性方面優於其他細胞類型。大多數細胞類型的 FID 值集中在 30 到 40 之間，顯示出這些生成圖像與真實圖像之間有著良好的相似度，具有較高的真實性。這些結果表明，生成模型在多數細胞類型上表現穩定，能夠產生真實性較高的圖像。

〔表 4. 生成影像多樣性指標結果〕

細胞種類	**CMMD**
Smudge Cells	0.0634
Promyelocyte	0.1591
NRBC	0.1467
Neutrophilic Myelocyte	0.1598
Neutrophilic Metamyelocyte	0.1491
Neutrophil Segment	0.1235
Monocyte	0.1220
Lymphocyte	0.1518
GIANT PLT	0.1661
Eosinophils	0.1392
Blast	0.1321
Basophils	0.0878
Band	0.1234
Atypical lymphocyte	0.1678

根據 CMMD 指標的結果，Smudge Cells 顯示出極佳的多樣性，CMMD 值為 0.0634，表現最為優異。大多數細胞類型的 CMMD 值也展現了穩定的多樣性，其中 Neutrophil Segment(0.1235)和 Monocyte(0.1220)顯示出較高的多樣性，顯示生成圖像的多樣性表現良好。這些結果表明，生成模型能夠在多數細胞類型上有效產生多樣化的圖像，並且在大部分情況下，生成圖像的多樣性都達到了令人滿意的水準。

6.1.2 生成影像之檢驗人員辨認結果

(a) 檢驗人員 1（年資：5 年以上）

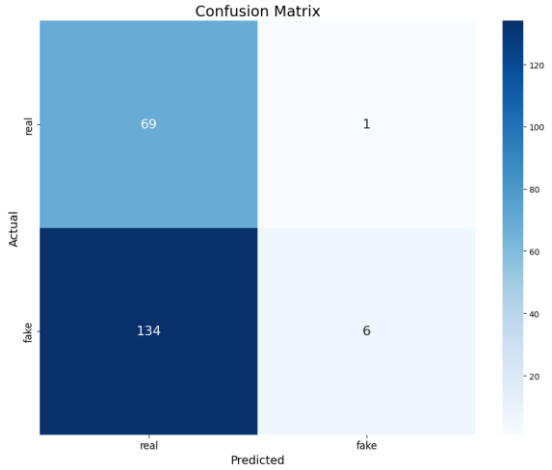


圖 9. 檢驗人員 1 辨別真偽資料數之混淆矩陣(confusion matrix)

(b) 檢驗人員 2（年資：5 年以上）

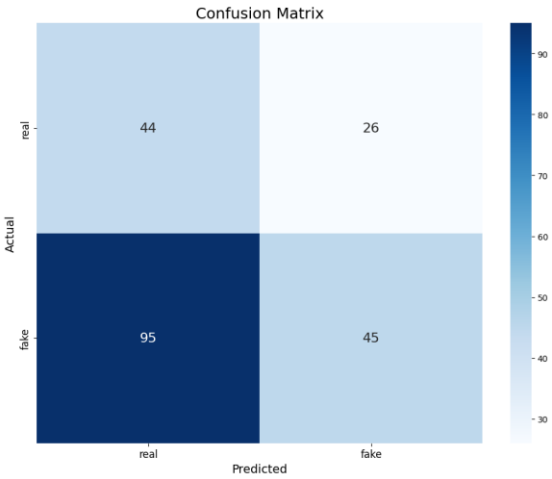


圖 10. 檢驗人員 2 辨別真偽資料數之混淆矩陣

(c) 檢驗人員 3（年資：5 年）

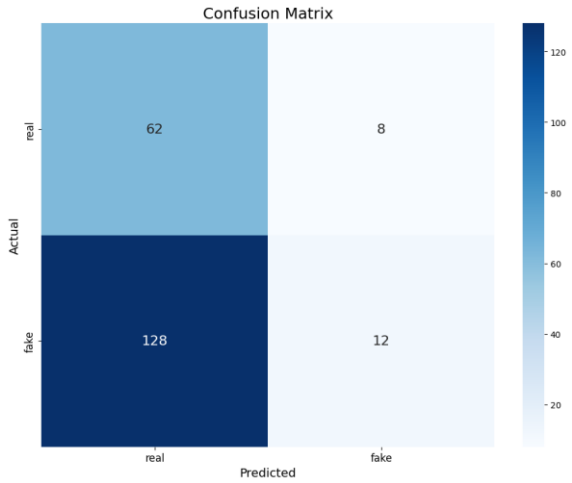
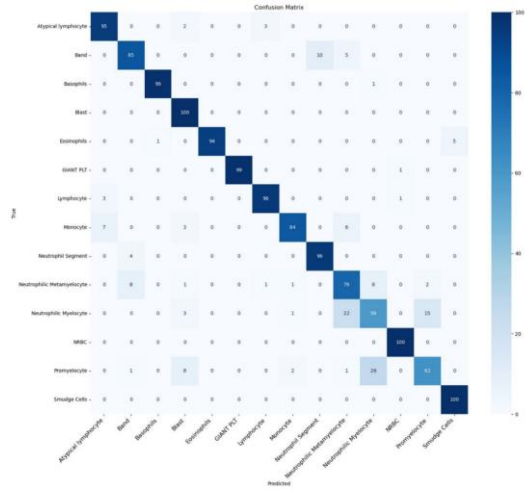


圖 11. 檢驗人員 3 辨別真偽資料數之混淆矩陣

圖 9、圖 10、圖 11 的 confusion matrix 皆顯示出，大多數以 Diffusion-GAN 生成的圖片，對於檢驗人員而言，與真實細胞圖片具高度相似性，即使是已獲得國家認證的細胞型態分類臨床專家證書人員也難以辨別。

6.2 模型分類效能與專業檢驗人員比對



6.2.1 CLIP 模型分類結果

圖 12. CLIP 模型在 14 類白血球細胞分類任務中，Top-1 預測結果之混淆矩陣

	precision	recall	f1-score	support
Atypical lymphocyte	0.9490	0.9300	0.9394	100
Band	0.9186	0.7900	0.8495	100
Basophils	1.0000	0.9800	0.9899	100
Blast	0.8547	1.0000	0.9217	100
Eosinophils	0.9706	0.9900	0.9802	100
GIANT PLT	0.9900	0.9900	0.9900	100
Lymphocyte	0.9700	0.9700	0.9700	100
Monocyte	0.9663	0.8600	0.9101	100
Neutrophil Segment	0.8545	0.9400	0.8952	100
Neutrophilic Metamyelocyte	0.7130	0.7700	0.7404	100
Neutrophilic Myelocyte	0.6134	0.7300	0.6667	100
NRBC	0.9804	1.0000	0.9901	100
Promyelocyte	0.8235	0.5600	0.6667	100
Smudge Cells	0.9515	0.9800	0.9655	100
accuracy			0.8921	1400
macro avg	0.8968	0.8921	0.8911	1400
weighted avg	0.8968	0.8921	0.8911	1400

圖 13. CLIP Top-1 分類資料類別之分類報告(classification report)

6.2.2 檢驗人員對 14 類白血球細胞之分類結果

(a) 檢驗人員 1（年資：5 年以上）

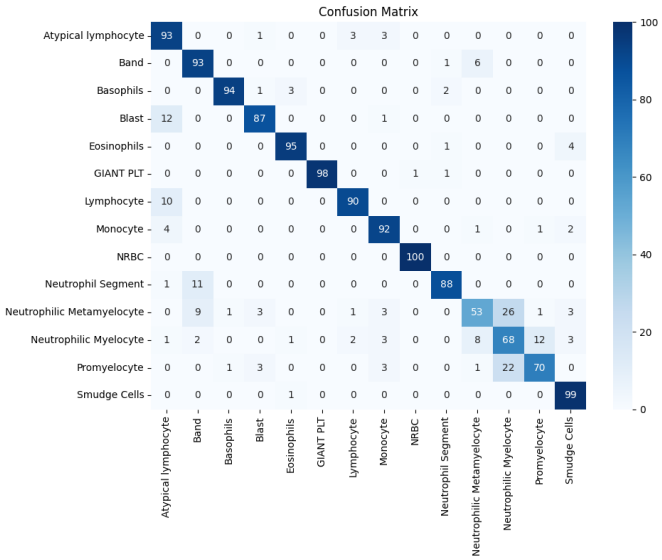


圖 14. 檢驗人員 1 辨別資料類別之混淆矩陣

Classification Report:

	precision	recall	f1-score	support
Atypical lymphocyte	0.77	0.93	0.84	100
Band	0.81	0.93	0.87	100
Basophils	0.98	0.94	0.96	100
Blast	0.92	0.87	0.89	100
Eosinophils	0.95	0.95	0.95	100
GIANT PLT	1.00	0.98	0.99	100
Lymphocyte	0.94	0.90	0.92	100
Monocyte	0.88	0.92	0.90	100
NRBC	0.99	1.00	1.00	100
Neutrophil Segment	0.95	0.88	0.91	100
Neutrophilic Metamyelocyte	0.77	0.53	0.63	100
Neutrophilic Myelocyte	0.59	0.68	0.63	100
Promyelocyte	0.83	0.70	0.76	100
Smudge Cells	0.89	0.99	0.94	100
accuracy			0.87	1400
macro avg	0.88	0.87	0.87	1400
weighted avg	0.88	0.87	0.87	1400

圖 15. 檢驗人員 1 辨別資料類別之分類報告

(b) 檢驗人員 2（年資：5 年以上）

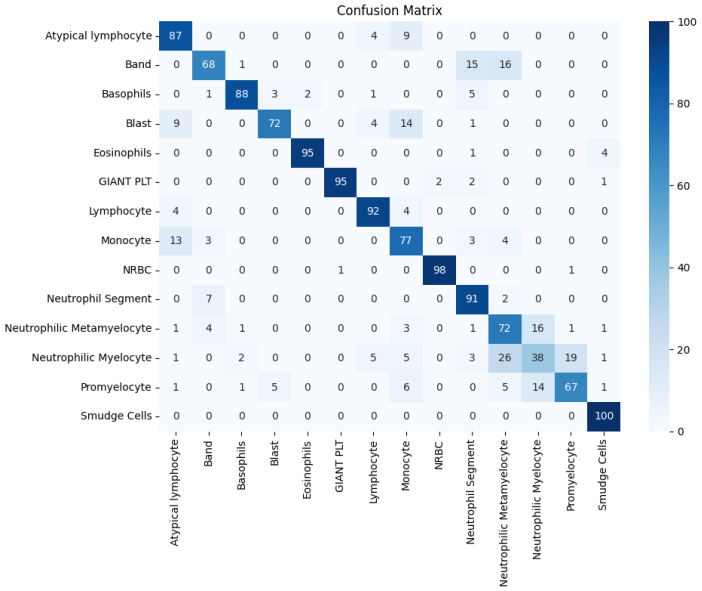


圖 16. 檢驗人員 2 辨別資料類別之混淆矩陣

Classification Report:

	precision	recall	f1-score	support
Atypical lymphocyte	0.75	0.87	0.81	100
Band	0.82	0.68	0.74	100
Basophils	0.95	0.88	0.91	100
Blast	0.90	0.72	0.80	100
Eosinophils	0.98	0.95	0.96	100
GIANT PLT	0.99	0.95	0.97	100
Lymphocyte	0.87	0.92	0.89	100
Monocyte	0.65	0.77	0.71	100
NRBC	0.98	0.98	0.98	100
Neutrophil Segment	0.75	0.91	0.82	100
Neutrophilic Metamyelocyte	0.58	0.72	0.64	100
Neutrophilic Myelocyte	0.56	0.38	0.45	100
Promyelocyte	0.76	0.67	0.71	100
Smudge Cells	0.93	1.00	0.96	100
accuracy			0.81	1400
macro avg	0.82	0.81	0.81	1400
weighted avg	0.82	0.81	0.81	1400

圖 17. 檢驗人員 2 辨別資料類別之分類報告

（c）檢驗人員 3（年資：5 年）

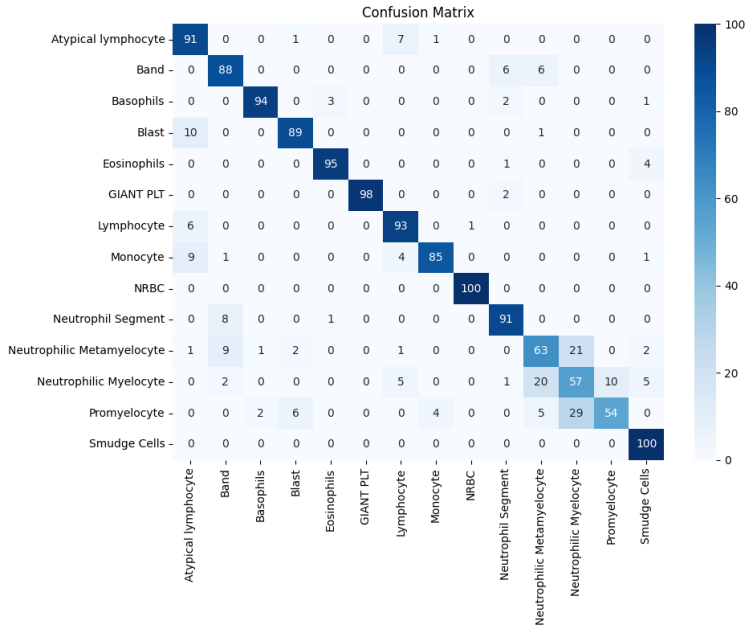


圖 18. 檢驗人員 3 辨別資料類別之混淆矩陣

Classification Report:

	precision	recall	f1-score	support
Atypical lymphocyte	0.78	0.91	0.84	100
Band	0.81	0.88	0.85	100
Basophils	0.97	0.94	0.95	100
Blast	0.91	0.89	0.90	100
Eosinophils	0.96	0.95	0.95	100
GIANT PLT	1.00	0.98	0.99	100
Lymphocyte	0.85	0.93	0.89	100
Monocyte	0.94	0.85	0.89	100
NRBC	0.99	1.00	1.00	100
Neutrophil Segment	0.88	0.91	0.90	100
Neutrophilic Metamyelocyte	0.66	0.63	0.65	100
Neutrophilic Myelocyte	0.53	0.57	0.55	100
Promyelocyte	0.84	0.54	0.66	100
Smudge Cells	0.88	1.00	0.94	100
accuracy			0.86	1400
macro avg	0.86	0.86	0.85	1400
weighted avg	0.86	0.86	0.85	1400

圖 19. 檢驗人員 3 辨別資料類別之分類報告

從圖 14 至圖 19 可看出，資深檢驗人員對於白血球細胞進行分類時，準確率(accuracy)分別為 87%, 81%, 86%，即使是最高的準確率也僅有約 87%，而此分類模型準確率可達 89%，此模型在對細胞進行分類時表現出極高的準確率，超越三位已獲得國家認證的細胞型態分類臨床專家證書人員的分辨能力，並展現出「資深檢驗人員」級別的專業水平。

6.3 LLAMA 模型生成敘述準確度

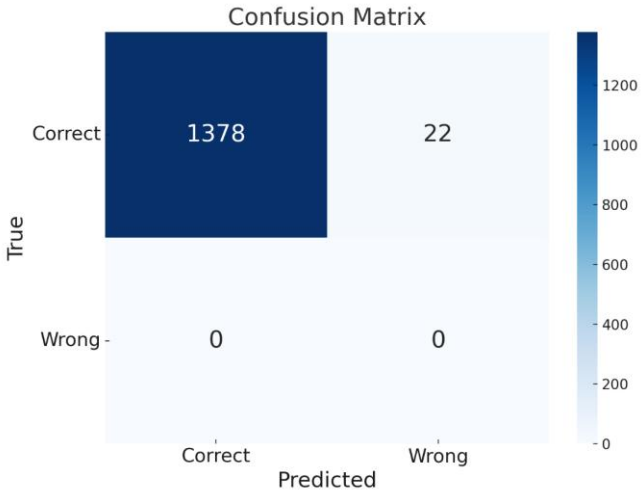


圖 20. 圖為模型對影像進行推論並生成對應文字之結果，文字描述與標準答案之 Soft Cosine Similarity 評估描述內容之語意相似度。分數若高於 0.7 則為正確。

	precision	recall	f1-score	support
Correct	1.0000	0.9843	0.9921	1400
Wrong	0.0000	0.0000	0.0000	0
accuracy			0.9843	1400
macro avg	0.5000	0.4921	0.4960	1400
weighted avg	1.0000	0.9843	0.9921	1400

圖 21. LLaMA 生成敘述準確率之分類報告

6.4 CLIP 結合 LLAMA 之推論結果

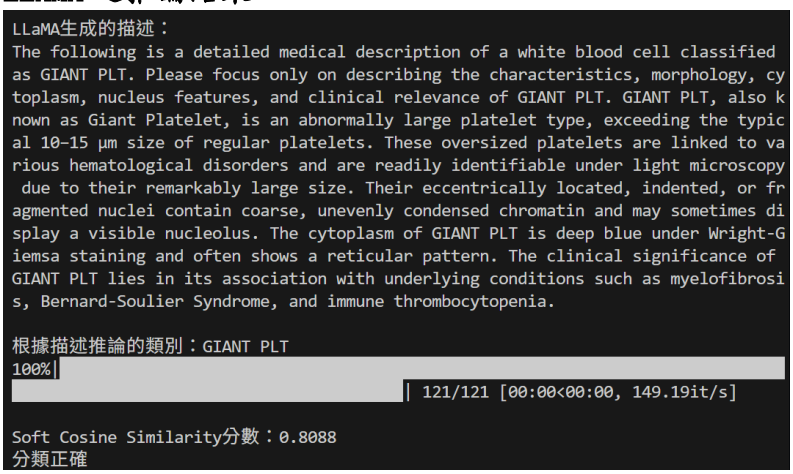


圖 22. 圖為模型對影像進行推論之結果，顯示 Top-1~3 分類且顯示分類準確率，並由 LLaMA 模型生成對應文字描述。文字描述與標準答案之 Soft Cosine Similarity 分數，評估描述內容之語意相似度。

6.5 提案優點與貢獻

本提案以 CLIP ViT-H/14 為基礎，針對白血球細胞分類進行微調，並結合大型語言模型 LLaMA-3-8B，建立一套分類準確率達到 89%以上且同時具備醫學描述生成功能的系統。

此外，為解決資料集中部分細胞類別樣本不足與細胞間特徵相似性高所造成的分類困難，採用 Diffusion-GAN 進行資料增補，有效提升資料多樣性與模型訓練效果。針對生成資料，亦透過多項指標及具國家認證之細胞型態分類臨床專家證書人員進行品質驗證，確保影像真實性。

同時，本提案應用 LoRA 技術對 LLaMA-3-8B 進行輕量化微調，使其能針對白血球細胞影像生成符合臨床標準的醫學描述文本，進一步提升系統的解釋能力與應用價值。

(a) 解決設備成本問題並促進偏鄉醫療發展

傳統細胞分類設備價格昂貴且維護成本高，不利於普及至資源有限地區。本提案系統僅需檢驗人員使用手機結合顯微鏡，拍照後連接到伺服器並上傳，如圖 25，將影像上傳至我們開發的 APP 後(可以透過手機拍攝或者相簿選擇)，如圖 24，即可進行即時細胞分類以及輸出分類依據，且分類準確率達到 89%以上。與傳統方法相比，本系

統建置成本大幅降低，且分類準確率超越三位獲得國家認證之細胞型態分類臨床專家人員的辨識能力。我們也將本研究成果的模型與衛生福利部桃園醫院檢驗科所租用的高價儀器 DI-60 進行比較，該設備具備六類白血球分類功能，因此我們亦讓模型針對相同六類細胞進行分類，並與其進行對照，結果如圖 23 的混淆矩陣所示，我們的模型在整體分類效能上已超越該設備。未來可推廣至偏鄉或基層醫療單位，節省人力與時間成本，分擔資深檢驗人員的工作的同時，輸出分類依據也能夠達到教育訓練的目的。

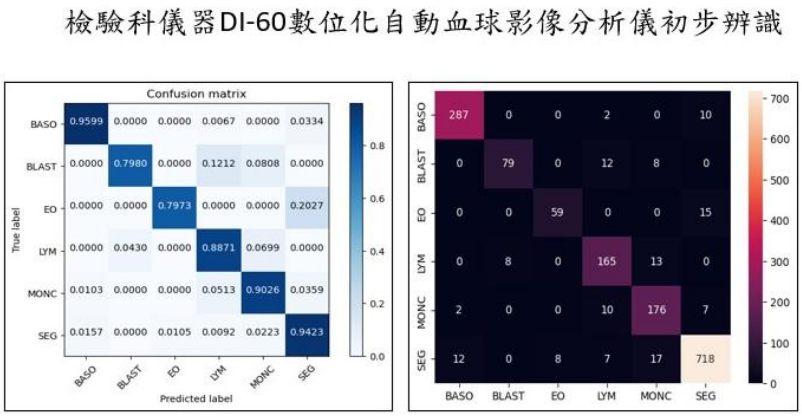


圖 23. 檢驗科儀器辨別六類白血球資料之混淆矩陣

(b) 協助資淺人員進行細胞分類與學習

本 APP 系統除提供影像分類結果外，亦引用跳脫了傳統僅依影像特徵分類的 Text as Hub 框架，如圖 24，透過將影像特徵轉換為文字描述，同步生成細胞醫學描述，有助於資淺檢驗人員或醫學生迅速掌握細胞特徵、分類原因，提升分類效率的同時也能夠進行學習。

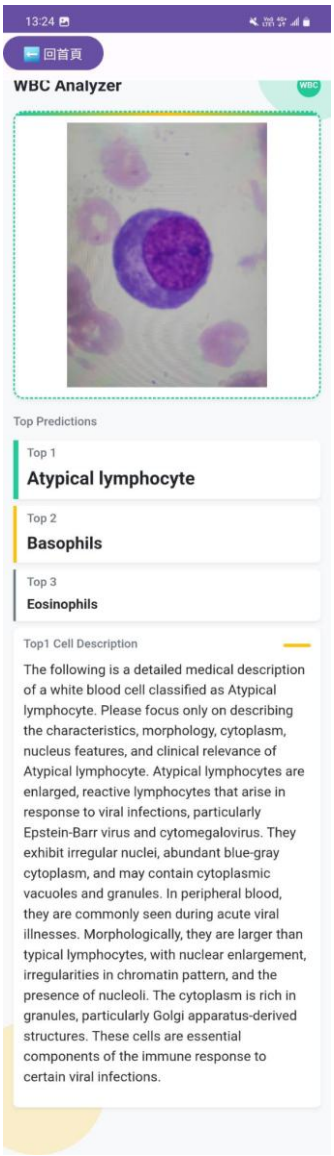


圖 24. 手機介面的長截圖(包括手機相機拍攝的細胞照片、分類結果 Top3 以及分類描述)

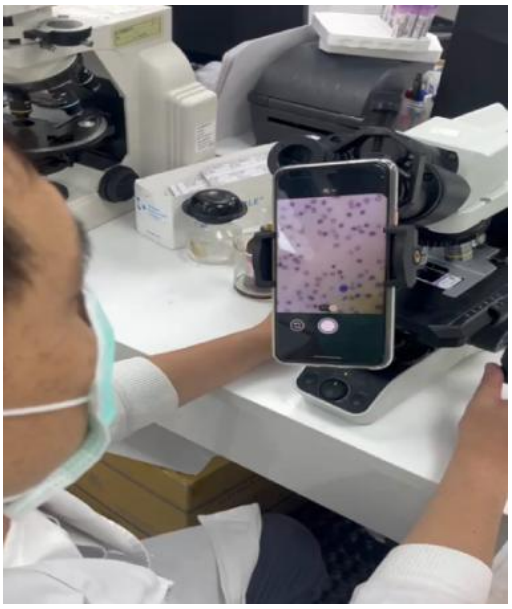


圖 25. 檢驗人員使用手機 APP 在顯微鏡進行細胞辨識

Chapter 7 誌謝

在此，我們要衷心感謝在這個研究過程中提供無私幫助與支持的所有人，沒有您們的協助，我們無法順利完成這項研究。

首先，非常感謝衛生福利部桃園醫院醫事檢驗科血液鏡檢組組長葉建宏先生提供的資料，特別是在整個資料收集過程中的大力支持，不僅提供了非常珍貴的數據，還找了多位權威級細胞型態學專家對資料進行了詳細複查，確保資料的準確性，為我們的研究工作節省了大量處理資料集的時間與精力。

其次，我們感謝吳昱壬學長，他在這項研究開始前，已測試過多種影像生成模型的效能。此經驗幫助我們在最短的時間內選擇最合適的模型，並讓我們可以更順利地進行訓練。

我們尤其要感謝謝瑞建教授在整個研究過程中提供的寶貴技術指導。研究初期，我們在 CLIP 模型的微調策略設計、影像特徵提取流程，以及 LLaMA 模型結合多模態推論時，曾面臨訓練穩定性不足、特徵串接錯誤與生成效果不佳等多方面技術挑戰。謝老師耐心指導我們釐清問題根源，並提供了具體的調整建議，如模型參數設定、訓練流程優化及錯誤排除的方法。透過謝老師的指導，我們得以有效提升模型效能，縮短了大量反覆試錯的時間，讓研究工作得以高效推進並順利完成。

我們也要誠摯感謝邱昭彰教授在文獻閱讀與研究過程中的悉心指導。每當我們在挑選相關文獻或閱讀內容時遇到困難，邱老師總能協助我們迅速掌握重點，並指引我們如何將論文中的理論實際應用於實驗之中。

最後，我們衷心感謝一路支持與陪伴我們的人們，無論是教授還是同學，正因為有你們每一份鼓勵與協助，才成就了這份成果。

Reference

[1] 醫療財團法人台灣血液基金會，（無年份），「血液基本介紹」，取自：https://www.blood.org.tw

[2] Supawit Vatathanavaro, Suchat Tungjitnob, Kitsuchart Pasupa, “White Blood Cell

Classification: A Comparison between VGG-16 and ResNet-50 Models," Faculty of Information Technology, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand.
[3] Wang, D., Hwang, M., Jiang, W.C., Ding, K., Chang, H.C., Hwang, K.S., "A deep learning method for counting white blood cells in bone marrow images"
[4] Ferdousi, J., Lincoln, S.I., Alom, M.K., Foysal, M., (2024), "A deep learning approach for white blood cells image generation and classification using SRGAN and VGG19"
[5] Barrera, K., Merino, A., Molina, A., Rodellar, J., (2023), "Automatic generation of artificial images of leukocytes and leukemic cells using generative adversarial networks (syntheticcellgan)"
[6] Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., Greenspan, H., (2018), "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification", Medical Image Computing and Computer Assisted Intervention – MICCAI 2018, Lecture Notes in Computer Science, Vol. 11070, pp. 206–214
[7] Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., Sutskever, I.,（2021），"Learning Transferable Visual Models From Natural Language Supervision", arXiv preprint arXiv:2103.00020
[8] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., (2021), "Attention Is All You Need", arXiv preprint arXiv:2304.0264
[9] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N., (2020), "An image is worth 16x16 words: Transformers for image recognition at scale", arXiv preprint arXiv:2010.11929
[10] Shaw, P., Uszkoreit, J., Vaswani, A., (2018), "Self-attention with relative position representations", arXiv preprint arXiv:1803.02155
[11] Gao, P., Zhu, M., Zhan, X., Dai, J., Lin, D., Qiao, Y., (2021), "CLIP-adapter: Better vision-language models with feature adapters", arXiv preprint arXiv:2110.04544
[12] Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., et al., (2023), "LLaMA: Open and Efficient Foundation Language Models", arXiv preprint arXiv:2302.13971
[13] Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Sylvain Gugger, Mariam Abdou, Roman Dettmers, Jared Casper, Younes Belkada, Armand Joulin, (2023), "Llama 2: Open Foundation and Fine-tuned Chat Models," arXiv preprint, arXiv:2307.09288.
[14] Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., de Laroussilhe, Q., Gesmundo, A., Attariyan, M., Gelly, S., (2019), "Parameter-efficient transfer learning for NLP", arXiv preprint arXiv:1902.0075
[15] Llama Team, AI@Meta, (2024), "The Llama 3 Herd of Models," arXiv preprint, arXiv:2407.21783.
[16] M. Yıldırım, A. Cinar, (2019), "Classification of White Blood Cells by Deep Learning Methods for Diagnosing Disease," Revue d'Intelligence Artificielle, Vol. 33, No. 5, pp. 319–326

附件 A、工作內容

一、工作環境

我們的實驗室位於 一館六樓的 1617 室，環境設備完善，老師的辦公室與實驗室是同一間，，有任何問題都能立即討論，整體溝通非常方便。

實驗室設有 VPN，因此我們在進行專題開發時，不需要每次都到現場。平常只要透過 VPN，就能遠端連線到實驗室的主機進行開發與測試。這讓我們在安排時間上相當彈性，也能避免因為硬體需求而受限於位置。

在使用的主機部分，主要是透過 VPN 連線到實驗室的深度學習設備，包括：

- RTX 4090（24GB VRAM）
- RTX 3090（24GB VRAM）
- RTX 2060（12GB VRAM）

這些主機負責模型訓練與推論，能滿足白血球影像分類與描述模型在效能上的需要。

在開發環境方面，我們使用 Python、Flask、PyTorch、HTML/CSS/JavaScript 進行後端與前端的製作，並搭配 SQLite 資料庫來處理預測結果與紀錄管理。

另外，我們也有建立 Line 群組。老師會在群組發布公告、指派任務，或提醒各組進度。如果有需要開遠端會議，也會利用 Line 或 Teams 簡單進行遠端討論。

二、工作內容

- ✓ 使用 PyTorch 訓練進行白血球影像分類與描述模型的訓練與調整。
- ✓ 開發 Flask API 串接模型並建置 SQLite 資料庫，完成前端網頁介面與 App 端 WebView 的整合。
- ✓ 使用 Android Studio 製作 App 介面，影像上傳流程與後端 API 的預測結果顯示。

附件 B、自我評估及心得感想

1.李秉霖:

在這次的專題研究中，我主要負責 CLIP 模型與 LLAMA 的設計。研究的核心是希望透過跨模態對比學習，將「白血球影像特徵」與「細胞類別文字描述」對齊於共享語義空間，藉此提升分類效能。同時，我們也結合 LLaMA，使系統不僅能完成分類，還能輸出臨床解釋性的診斷描述，協助資淺醫檢師在分類的同時理解診斷邏輯。

在研究過程中，我們面臨了許多挑戰。首先是資料不足的問題，部分稀有白血球類別樣本極少，導致模型難以有效學習。而為了解決這個問題，我們引入 Diffusion-GAN 的技術，生成與真實影像相近的白血球影像，來補足稀有白血球的影像。然而在實驗過程中，我們遭遇了訓練不穩定、程式崩潰、判別器過度擬合等狀況，需要不斷調整訓練流程與模型參數，才能讓生成影像達到可用的品質。

在模型選擇上，我一開始採用了 EfficientNetV2 進行分類實驗，雖然在常見的 6 類白血球分類上表現不錯，但當任務擴展到 14 類白血球時，準確率並未達到我們的預期。為了突破這個瓶頸，我深入查閱相關文獻，並發現 CLIP 架構中的 Vision Transformer (ViT) 具有更強大的特徵抽取與跨模態對齊能力，並在經過多次實驗與調整，我們最終成功將分類準確率提升至 89%，不僅超越了具有專業認證的醫檢師水準，甚至優於現今市面上常見的白血球數位影像處理系統 DI-60。

LLAMA 的設計，主要負責將 CLIP 提取出的影像特徵，透過 Text-as-Hub 轉換成語義嵌入，再輸入 LLaMA 進行報告生成。為了讓輸出的描述更符合臨床需求，我設計了專屬的醫學提示詞，並透過 LoRA 輕量化微調，使模型能夠生成具專業性且條理清楚的診斷依據。這樣的設計不僅提升了系統的臨床可解釋性，也讓模型具備教育輔助價值，能幫助年資較淺的醫檢師理解模型的判斷依據。

完成研究後，我們有將成果參加第十四屆全國大專院校 AI 智動化設備創作獎，並榮獲第二名。能在全國性競賽中獲獎，對我而言是一個極大的肯定。這不僅證明我們的研究成果具備實用價值，也讓我學會如何將學術研究轉化為能被外界理解的應用方案。比賽過程中，與評審與業界專家的交流，讓我更加體會到醫療 AI 的落地應用前景，也更加堅定了我繼續投入此領域的決心

2.游凱甯:

在這次專題研究中，讓我感到最有收穫的並不是最終的成果，而是逐步理解模型特性、反覆驗證想法的過程。專題初期，我們並沒有預期會把 CLIP、LLaMA 與 Text-as-Hub 結合得如此深入；但隨著研究慢慢推展，我開始投入於跨模態模型如何同時理解白血球影像與文字描述這件事。當時的核心想法相當單純：只要能讓影像特徵與語意資訊在同一空間中對齊，分類表現必定能有所提升。也因此，我負責了整體 CLIP 訓練流程的設計與調整，從資料處理到模型架構修改，都必須不斷反覆確認與驗證。

真正具挑戰性的部分，是訓練過程中接連出現的各種不確定性。模型偶爾會突然崩潰、生成器與判別器互相牽制、loss 一度停滯不動，甚至還遇過過度擬合導致結果失真。為了找出每個問題背後的原因，我花了大量時間調整參數、更改架構，這些經驗雖然耗時，也有些挫折，但卻讓我第一次真正理解「讓模型穩定」比「讓模型變強」更需要技術與耐心。

模型的分類表現也不是一路順利。我們曾遇到長時間無法突破的準確率瓶頸，怎麼調都不見起色。後來透過重新規劃資料分布、調整正負樣本策略，以及改善跨模態對齊方式，結果才突然有了明顯改善，最終達到 89% 的準確率。後續比對才發現，這個表現甚至高於一般醫檢師以及 DI-60 系統，對我們而言是相當意外也令人鼓舞的里程碑。

在完成 CLIP 之後，我們接著讓 LLaMA 承接影像的語意嵌入，使系統不只會分類，更能產生診斷式的文字說明。為了讓描述更貼近臨床語氣，我重新整理醫學術語並設計提示詞模板，再以 LoRA 進行微調。這讓模型能在輸出分類結果的同時，清楚說明它做出判斷的依據，對於初學者或資淺醫檢師而言，是很直接的學習輔助。我們也將成果送往全國性競賽，最後獲得第二名。對我而言，這份肯定比獎狀本身更具意義，因為它代表研究不只是停留在程式碼或論文中，而是能被他人理解並看見價值。

回顧整個專題，我重新認識了「訓練模型」這件事。它不是單純的超參數組合或技巧堆疊，而是一連串試驗、推敲與理解的累積。在一次次的調整與修正中，學到如何讓系統真正變得更成熟、更可靠。

3.廖奕玟:

在本次專題研究中，我從資料處理、模型訓練、生成式影像技術到語言模型的整合，完整體驗了一個醫療 AI 系統從構思、模型選擇、驗證到成果展示的完整過程。這項研究不僅加深我對深度學習與多模態模型的理解，也使我更深刻體會到，人工智慧在醫療領域應用時，必須兼具技術精確性與臨床實用性，否則即使模型效能再好，也無法真正被使用者採納或信任。

在研究初期，我們觀察到醫療設備在偏鄉醫療資源不足情況下的侷限性，因此希望透過多模態人工智慧技術，讓診斷輔助工具不再依賴昂貴儀器，而能以更平易近人的方式協助臨床人員。

我們的資料來源來自衛生福利部桃園醫院檢驗科，涵蓋 14 類白血球影像，但其中不乏樣本極低的罕見細胞，造成訓練時的資料不平衡問題。面對這項挑戰，我們採用了 GAN 進行資料增補，使每一類別樣本達到較均衡的數量，並以 FID 與 CMMD 評估生成的影像品質，同時邀請臨床專家進行盲測，結果顯示生成影像與真實影像極為接近，整個評估過程都非常嚴謹，這讓我親身體會到資料品質的重要性。

在模型設計方面，我們並未採用傳統 CNN，而是選擇 CLIP 結合 Text as Hub 的概念，利用影像與語言共同學習的方式進行分類，並加入自訂 MLP 層以進行微調。實驗結果顯示我們的系統分類準確率達到 89%，甚至超越三位具國家臨床認證的檢驗人員，這個結果讓我非常驚訝，也讓我重新思考 AI 不僅可以是輔助工具，甚至能成為臨床決策的重要參考依據。此外，我們也結合 LLaMA 生成敘述並進行語意可信度評估，使系統能夠針對分類結果去做解釋，這對臨床教學、資淺人員訓練都具有實際價值。

透過本研究，我重新認識到醫療 AI 的核心不僅在於追求高準確率，更必須同時兼具可靠性、可解釋性、實際落地性與倫理安全。在這項研究中，我的收穫不只成果本身，而是過程中逐步建立的問題分析能力與研究習慣。專題進行期間遇到的困難並非單一領域，而是涵蓋資料、模型、實驗規劃、溝通與驗證，這讓我重新理解研究不是線性流程，而是持續假設、修正與驗證的反覆循環。這段過程讓我學會在不確定性中保持判斷能力，也更明白投入研究時，耐心、紀錄、反思與反覆實驗的重要性。

4.陳薇珊:

在這次專題研究中，我完整經歷了一個醫療 AI 系統從資料整理、模型設計、生成式影像製作到多模態整合的整體流程。這段過程讓我深刻理解到，醫療人工智慧並不是只要模型分數等各類性能指標漂亮就好，而是必須同時兼顧臨床需求、資料品質與實際使用情境。只有在技術與臨床之間取得平衡，AI 才可能真正被醫療人員採納並發揮價值。

研究一開始，我們注意到偏鄉醫療資源有限，而許多檢測設備如 DI-60 昂貴又需要專業技師操作，因此希望透過多模態 AI 方式提供更普及的輔助診斷工具。這個動機讓我更明白技術發展的目的應該是「解決真實問題」，而非純粹追求更高的指標。本研究使用桃園醫院檢驗科的 14 類白血球影像，其中不乏樣本極少的罕見類別，造成訓練資料明顯不均衡。為了提升模型的穩定性，我們以 Diffusion-Projected GAN 補足稀少類別，並透過 FID、CMMD 等多項指標與臨床盲測多重驗證生成品質。當看到專家實際分辨不出生成與真實影像的差異時，我第一次真切感受到「資料品質」對整個 AI 系統的重要性，甚至遠大於模型結構本身。

在模型設計上，初期嘗試了 EfficientNetV2 與 Mobile ViT ，後期沒有沿用傳統的 CNN，而是選擇採用 CLIP 並加入 Text as Hub 的概念，將影像特徵與文字描述共同作為分類依據。同時，我們以自訂 MLP 進行微調，使模型能更好地理解醫療語意。最終系統的分類表現達到 89%，甚至超越臨床檢驗師的平均表現，這讓我重新思考 AI 在臨床中的角色，它不只是輔助工具，也可能成為醫療決策的重要參考。此外，我們利用 LLaMA 生成診斷敘述，並評估語意準確性，使模型不只會預測，也能「說明理由」，提升其在教學與臨床上的可用性。

整個研究中，我逐漸理解醫療 AI 的價值不只在於提高辨識率，更在於滿足「可靠性、可解釋性、可落地、具倫理性」等多層需求。這個專題帶給我的最大成長，不是模型跑得多好，而是我在面對資料混亂、模型不收斂、實驗設計反覆修改、與臨床人員溝通等挑戰時，累積出的研究思維與問題分析能力。我也更明白研究並不是直線前進，而是持續假設、驗證、修正的循環過程。

這段經驗讓我建立了在不確定環境中保持冷靜、耐心與紀錄習慣的能力，也更確信自己想在 AI 與醫療的交叉領域繼續深耕。

5. 江昀芮

在這次專題研究中，我主要負責後端 API、資料庫設計，以及前後端系統整合，確保模型輸出能順利被系統接收並呈現給使用者。雖然模型訓練不是我主要的工作，但透過串接的過程，我對模型的輸入、輸出與運作方式有了更完整的理解，也因此對整個系統的架構與流程有更清楚的掌握。在這過程中，我逐漸理解醫療 AI 系統「落地應用」的重要性：再高的模型準確率，如果無法被臨床人員方便使用，也難以真正發揮價值。

在開發過程中，我主要著重於 App 與 Web 介面的前端設計與整合，確保檢驗人員操作流程直覺、順暢，並能即時上傳與記錄填寫資料。雖然在前端 UI 的設計過程中遇到不少挑戰，例如程式 debug、資料傳遞異常，以及 API 與模型串接的問題，但透過查找資料、反覆測試與調整，我逐漸掌握了系統整合與問題排查的技巧。這也讓我更熟練於 Flask API 的應用、Web 前端技術，以及整體系統的開發流程。

此外，我們透過系統收集到的影像資料，也可作為未來模型訓練的 dataset，持續精進模型效能，讓系統的準確率與穩定性不斷提升。最讓我印象深刻的是，我們有到醫院進行實際測試，看到系統真的能協助檢驗流程、對偏鄉資源不足的醫療單位可能有幫助時，讓我深刻體會到專題不只是課堂作業，而是具有真正應用價值的專案。透過這些實際操作，也讓我更理解資料收集、模型優化與系統落地的重要性，為後續改進提供了寶貴的參考。

整體而言，這次專題讓我在後端整合、前端介面設計、系統串接以及落地應用等方面都有了全面的體驗，也累積了大量實務經驗。我學會了如何在開發過程中保持細心與耐心，並理解到技術與使用者需求之間的平衡。這些經驗不僅提升了我的程式能力，也加深了我對醫療 AI 系統從設計、開發到落地應用流程的理解，為未來相關領域的研究與開發打下了扎實基礎。