

國立雲林科技大學工業工程與管理系

碩士論文

Department of Industrial Engineering and Management

National Yunlin University of Science & Technology

Master Thesis

基於深度學習之鼻咽癌腫塊辨識

Recognition of Nasopharyngeal Carcinoma

Discrimination from Endoscopic Images Based on Deep
Learning Models

吳文成

Wen-Cheng Wu

指導教授：陳奕中 博士

Advisor: Yi-Chung Chen, Ph.D.

中華民國 110 年 6 月

June 2021

摘要

鼻咽癌(Nasopharyngeal Carcinoma, NPC)為國人目前常見的癌症疾病之一，但是其前期很難有明顯徵兆顯示是否有罹患此疾病，需依靠醫師藉由診斷工具診斷其疾病，而在所有鼻咽癌之診斷工具中內視鏡診斷正確率相對較低，若是長時間的診斷，對於病人及醫師皆不是樂觀的事情。在過往內視鏡影像辨識中，由於隨機性較大且可能受到燈光等雜訊干擾，且僅靠單一模型進行辨識，容易導致無法立即校正的問題，因此為了解決以上問題提出了基於深度學習之集成學習模型的方案，透過影像前處理來去除內視鏡影像內之雜訊，模型不受雜訊干擾而使模型誤判，考慮內視鏡環境，將原始色影像轉為不同色彩空間及不同亮度之影像使得影像有物體遠近的視覺效果及適應於內視鏡環境之亮度，最後透過集成學習模型輸出結果。在集成學習輸入的部分，以原始色、不同亮度、不同色彩空間之影像各別輸入至卷積類神經網路(Convolutional Neural Network, CNN)模型而 You Only Look Once(YOLO)模型的部份僅以原始色進行模型訓練，接著以不同色彩空間、不同亮度、YOLO 集成、YOLO 與 CNN 之排列組合的各別輸出結果作為各別集成之輸入，最後實驗結果集成後的效果遠大於單一模型，並且在準確度可以高達 97%，證明該方案能有效的對內視鏡鼻咽腫塊影像進行辨識。

關鍵字：鼻咽癌、深度學習、內視鏡、CNN、YOLO、集成學習

Abstract

Nasopharyngeal carcinoma (NPC) is common in Taiwan; however, disease detection in the early stages can be difficult without obvious clinical indications. Existing endoscopic image recognition schemes are hindered by the random nature of the data and interference from noise. In the current study, we developed an ensemble deep-learning model in which noise is removed via pre-processing and the original color image is converted to a different color space to enable the detection of object distance despite the confined luminance characteristics of the endoscopic environment. Images with the original color as well as those using different color spaces and different brightness values are learned using the You Only Look Once (YOLO) convolutional neural network (CNN) model for training. The results of the ensemble learning model are far superior to those obtained using a single model, achieving accuracy of up to 97%. Experiment results demonstrate the efficacy of the proposed system in identifying nasopharyngeal lumps in endoscopic images.

Keywords : NPC、Deep Learning、Endoscopic、CNN、YOLO、Ensemble Learning

誌謝

首先要特別感謝指導老師奕中老師，給予我非常多的指導，讓我對人工智慧一塊領域有更深入的了解，也因此喜歡上這個學術上的知識，並且也給予我很多機會與廠商、醫院合作，讓原本不是那麼會與他人談報告的我有了很顯著的進步，並且在做案子時能讓我對於程式的部分能夠更精進。

再來要感謝 S 醫院盧醫師及汪特助給予研究資料的協助，使我能夠很順利的完成此研究，其中特別感謝盧醫師那麼有耐心的標記每一張影像及給予我很多鼓勵及祝福。

特別感謝實驗室的博班學長姐東錡學長、映均學姐、富正大哥、豪哥，學長姐們在碩一升碩二的暑假非常有耐心的聽我報告，使我在報告上有著大幅的進步，在每一次報告的時候給予我很多報告上的建議，並且還會分享學術上的知識，使我對人工智慧有更加不一樣的 understanding。

謝謝實驗室的同學及同屆碩班的同學，彼此互相砥礪、幫助，使我在做研究這條路上有了莫大的動力，讓我覺得研究所的生活每一天都過得非常豐富。

目錄

摘要.....	i
Abstract.....	ii
誌謝.....	iii
目錄.....	iv
表目錄.....	vi
圖目錄.....	viii
第一章 緒論.....	1
1.1 研究背景與動機.....	1
1.2 研究目的與範圍.....	3
1.3 研究流程.....	4
第二章 文獻回顧.....	6
2.1 基於機器學習影像處理及辨識方法.....	6
2.2 基於深度學習影像處理方法介紹.....	12
2.3 集成式學習在影像中的處理方式.....	17
2.4 小結.....	19
第三章 資料集與初步整理.....	20
3.1 紙本資料.....	20
3.2 CNN 內視鏡鼻咽部位資料.....	20
第四章 基於 CNN 模型之做法.....	22
4.1 擷取 ROI.....	23
4.2 強光雜訊去除.....	23
4.3 影像色彩空間及亮度轉換.....	24
4.4 資料增強設計.....	25
4.5 建立基礎影像辨識之 CNN 模型.....	25

4.5.1	CNN 集成學習之模型選擇	25
4.5.2	應用 CNN 之 Inception-ResNet 架構辨識內視鏡影像	27
4.5.3	CNN 模型訓練	31
4.5.4	集成學習框架	33
4.5.5	績效衡量	34
第五章	基於 YOLO 模型之做法	36
5.1	YOLO 內視鏡影像前處理	36
5.2	建立基礎影像辨識之 YOLO 模型	38
5.2.1	應用 YOLO 辨識內視鏡影像	38
5.2.2	YOLO 模型訓練	40
5.2.3	集成學習框架	42
5.2.4	績效衡量	42
第六章	實驗模擬	44
6.1	資料集與實驗參數介紹	44
6.2	CNN、YOLO 模型之有效性	45
6.3	模型可視化分析探討	52
6.4	集成學習之模型選擇	55
6.5	模型之績效與評估	59
6.6	模型質化合理性說明	61
第七章	結論與未來研究	63
參考文獻	64

表目錄

表 1 基於機器學習影像處理及辨識相關文獻.....	11
表 2 基於深度學習之醫學影像辨識之相關文獻.....	17
表 3 基於集成學習之醫學影像辨識之相關文獻.....	19
表 4 CNN 內視鏡鼻咽部位資料集	20
表 5 混淆矩陣.....	34
表 6 YOLO 內視鏡鼻咽部位標籤.....	37
表 7 影像轉換之參數及訓練之模型.....	45
表 8 各模型訓練集混淆矩陣.....	46
表 9 各模型測試集混淆矩陣.....	48
表 10 各模型訓練集實驗結果.....	50
表 11 各模型測試集實驗結果.....	51
表 12 各模型交叉驗證實驗結果.....	51
表 13 不同色彩空間集成學習分類器之交叉驗證實驗結果.....	56
表 14 不同亮度集成學習分類器之交叉驗證實驗結果.....	57
表 15 YOLO 集成學習分類器之交叉驗證實驗結果	57
表 16 排列組合 C_3^{16} 集成學習實驗結果	58
表 17 排列組合 C_4^{16} 集成學習實驗結果	58
表 18 排列組合 C_5^{16} 集成學習實驗結果	58
表 19 排列組合特徵計數表.....	58
表 20 各別集成學習訓練集混淆矩陣.....	59
表 21 各別集成學習測試集混淆矩陣.....	60
表 22 各模型訓練集實驗結果.....	60
表 23 各模型測試集實驗結果.....	60

表 24 集成學習交叉驗證實驗結果.....	61
表 25 正確辨識之集成學習輸入端特徵計數結果.....	61
表 26 錯誤辨識之集成學習輸入端特徵計數結果.....	62



圖目錄

圖 1 (a)CT[16]、(b)MRI[33]、(c)內視鏡影像[3]	2
圖 2 研究流程.....	5
圖 3 基於機器學習影像處理及辨識流程圖.....	6
圖 4 LBP 流程圖	7
圖 5 HOG 流程圖[15]	8
圖 6 SIFT 流程圖[40].....	9
圖 7 SVM 示意圖	10
圖 8 KNN 示意圖	11
圖 9 深度學習之影像辨識流程圖.....	12
圖 10 CNN 之示意圖	13
圖 11 YOLO 模型概念圖[49]	14
圖 12 IoU 視覺化示意圖	16
圖 13 錨點示意圖.....	16
圖 14 Bagging 範例圖	17
圖 15 Boosting 範例圖	18
圖 16 Stacking 範例圖.....	18
圖 17 鼻咽陰性範例圖.....	21
圖 18 鼻咽陽性範例圖.....	21
圖 19 CNN 研究流程圖	22
圖 20 取出 ROI 區塊流程圖.....	23
圖 21 強光雜訊去除流程圖.....	23
圖 22 HSV 圓柱坐標系.....	24
圖 23 原始色影像轉不同色彩空間示意圖.....	24
圖 24 原始色影像轉不同亮度示意圖.....	24

圖 25 資料增強示意圖.....	25
圖 26 CNN 不同色彩空間影像集成學習	26
圖 27 CNN 不同亮度影像集成學習	26
圖 28 CNN 混和式集成學習	26
圖 29 卷積過程示意圖.....	27
圖 30 池化過程示意圖.....	27
圖 31 全連接層示意圖.....	27
圖 32 Inception-ResNet 網路架構示意圖[56].....	28
圖 33 Stem 示意圖[56].....	29
圖 34 ResNet 示意圖[27]	30
圖 35 Inception-Resnet 各模組示意圖[56].....	30
圖 36 Reduction 模組示意圖[56]	30
圖 37 模型訓練過程.....	31
圖 38 各影像輸入模型示意圖.....	31
圖 39 ReLU 函數.....	32
圖 40 集成學習之概念圖.....	33
圖 41 YOLO 研究流程圖.....	36
圖 42 鼻咽標籤陰性範例圖.....	37
圖 43 鼻咽標籤陽性範例圖.....	37
圖 44 YOLOv3 網路架構圖[51].....	38
圖 45 FPN 架構示意圖[39].....	39
圖 46 YOLOv3 訓練過程.....	40
圖 47 影像輸入模型示意圖.....	40
圖 48 Leaky ReLU 函數.....	41
圖 49 FPN 輸出結果示意圖	41
圖 51 不同亮度示意圖(a)參數 0.8、(b) 參數 1.2、(c) 參數 1.8.....	45

圖 52 (a)CNN 之 Grad-CAM 陽性視覺化結果(b)CNN 之 Grad-CAM 陰性視覺化結果.....	53
圖 53 (a)YOLO 陽性視覺化結果(b)YOLO 陰性視覺化結果	54
圖 54 (a) 真實資料為陽性被辨識成陰性 CNN 之 Grad-CAM 視覺化結果(b) 真實資料為陽性被辨識成陰性 YOLO 視覺化結果	55
圖 55 (a) 真實資料為陰性被辨識成陽性 Grad-CAM 視覺化結果(b) 真實資料為陰性被辨識成陽性 YOLO 視覺化結果	55

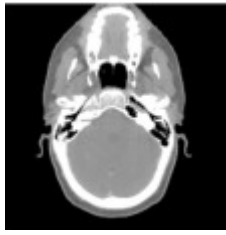


第一章 緒論

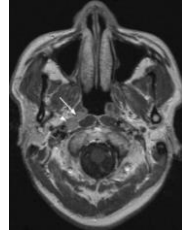
1.1 研究背景與動機

鼻咽癌為國人目前常見的癌症疾病之一，根據台灣三軍醫院資料顯示[1]，台灣每十萬人中罹患此癌症的人數為 7.7 人，男性較女性易患此疾病，其比例大約 3：1，而近年來，根據國內數據統計，每年大約有 1200 人罹患此癌症，其中此癌症為台灣男性十大癌症中的第七位，且多數人在罹患此疾病的前期很難有明顯徵兆顯示是否有罹患此疾病，僅少數人有鼻涕帶血、鼻塞、單側頭痛等症狀[1][2]，因此需要經過醫師的診斷，來判斷出是否有鼻咽癌之病徵，而以往診斷鼻咽癌的方式採用電腦斷層掃描攝影(Computed Tomography, CT)、磁力共振成像(Magnetic Resonance Imaging, MRI)、以及內視鏡拍下鼻咽部位之影像，接著醫師將從多張影像中分析是否有鼻咽癌之病徵[42]。

傳統的診斷鼻咽腫塊的流程為醫師先利用內視鏡進行檢查，將內視鏡深入鼻咽部位，接著針對鼻咽部位擷取多張影像，最後依序對這些擷取的影像進行診斷是否有鼻咽腫塊之病徵，若是有明確的鼻咽癌腫塊病徵，醫師將判斷此病患為具有鼻咽癌陽性腫塊，相反的，若是病患並沒有鼻咽癌腫塊之特徵，則將判斷此病患為陰性腫塊，但是在檢測鼻咽腫塊的流程會發生非常嚴重的困難點，根據 King 等人[33]統計結果顯示，內視鏡相較於 MRI 及 CT，鼻咽癌內視鏡診斷方式於臨床診斷中準確率相對較低，其原因為 MRI 及 CT 的影像較為固定，相較於內視鏡，內視鏡會依據醫師操作手法使得影像的隨機性較大，進而可能忽略鼻咽癌病徵，MRI、CT、內視鏡影像如圖 1 所示。在這種困難的狀況下，對於醫師來說，醫師在一整天當中需要診斷數百位病人，若是每一位病人都需要醫師依據數多張影像來判斷鼻咽腫塊是否為陰、陽性，使得整個診斷過程效率過低，且長時間下來，醫師的疲勞度可能會有較大的影響，在疲勞度越高的情況下，不管是對於病人或是醫師而言，皆不是一件樂觀的事情，未診斷出病人之病徵的風險可能會增加，進而造成醫療的品質下降；對於診斷鼻咽，由於每一位醫師在內視鏡操作手法上的不同，若是非主治醫師在診斷該影像時，需要根據主治醫師在內視鏡操作手法的角度去設想，最後會導致於非主治醫師在診斷的時間增加，疲勞度也會隨之增加。



CT(a)



MRI(b)



內視鏡(c)

圖 1 (a)CT[16]、(b)MRI[33]、(c)內視鏡影像[3]

過往鼻咽癌腫塊辨識的研究大多以 MRI、CT 為主，鮮少以內視鏡作為研究，其原因也是因為隨機性較大因素，使內視鏡影像於辨識中較為困難，此外內視鏡於拍攝中會由於人體中黑暗的環境必須要依靠燈光照明的輔助來幫助鏡頭能擷取下鼻咽部位之影像，但是在照明的過程當中可能會因為燈光照明到鼻咽部位進而產生出燈光反射之強光雜訊，這些強光雜訊可能會造成模型在訓練的過程當中學習到這些雜訊的特徵，進而使在模型的判斷上發生錯誤，同時，於內視鏡擷取影像中也會發生另一個常見的問題，人體產生之分泌物(如：鼻涕等。)於鼻咽部位中，在訓練模型時，亦會造成模型學習到分泌物及相關之不必要的特徵，進而導致模型判斷上有所誤差。

有鑑於上述的原因，因此學術界興起一波以影像處理及機器學習的方式協助醫師進行鼻咽癌的想法，不過這個想法仍然在萌芽當中，因為鼻咽癌辨識較為困難，因此相關研究較少，以下我們先回顧過往一些與醫學影像處理有關之方法，Korkmaz 與 Binol[34]於 2018 年利用局部二值模式(Local Binary Patterns, LBP)、方向梯度直方圖(Histogram of Oriented Gradients, HOG)先對胃癌影像擷取特徵，再由機器學習中隨機森林進行分類，找出是否有胃癌的徵兆；De Sampaio 等人[18]於 2015 年應用 LBP 對 X 光之乳癌影像擷取特徵並應用支援向量機(Support Vector Machine, SVM)進行分類，辨識是否有乳癌病徵；Alfed 與 Khelifi[6]於 2017 年藉由 HOG 與皮膚鏡影像中皮膚之紋理、顏色特徵結合，最後由 SVM 查找是否有皮膚癌病徵；Yuan 等人[63]於 2014 年使用基於傳統影像處理尺度不變特徵轉換(Scale-Invariant Feature Transform, SIFT)進行特徵擷取接著以 SVM 進行分類；Shiji 等人[54]於 2017 年採用 SIFT 並藉由詞袋模型(Bag of Features, BoF)找尋超音波

之乳癌影像之特徵，並由 SVM 分類辨識乳癌之陰、陽性；Yu 等人[62]於 2020 年透過 LBP、HOG、SIFT 對子宮頸癌影像擷取特徵，再由 KNN 及 SVM 分類其陰、陽性。

近年來隨著大數據技術的崛起，過去於影像辨識研究當中，Yu 等人[62]使用基於傳統影像處理 SIFT、HOG、LBP 進行特徵擷取接著以 SVM、KNN 進行分類，並與基於深度學習之 CNN 所擷取之特徵進行分類的比較，最終 CNN 的績效明顯高於基於傳統影像處理方法，其原因在於，基於傳統影像處理方法擷取特徵的方式 CNN 透過卷積的方式就能夠做到了，且 CNN 使用更多卷積核來擷取更多特徵值，使深度學習之 CNN 模型能夠比基於傳統影像辨識方法擁有更好的分類結果。而在 CNN 模型的使用中並不是完美的，從過往的研究可以發現 CNN 模型中僅依靠單一的模型進行辨識，導致於經過模型辨識後有無法修正的情況發生，使模型在一些影像的辨識上都是誤判的情況發生。

1.2 研究目的與範圍

從過往的研究發現以往的方法總共有幾個缺點，並將其缺點整理出以下問題，問題有內視鏡影像中雜訊、內視鏡特徵擷取不精確且隨機性大、僅參考單一辨識模型忽略其他重要分類模型，針對以上問題提出以下解決方案，首先於內視鏡影像中雜訊問題中，使用影像前處理針對內視鏡影像雜訊去除，以便於避免深度學習模型學習雜訊特徵，進而造成模型準確率下降，因此本研究透過影像前處理方法去除內視鏡影像中雜訊，使模型能學習內視鏡重要之特徵，接著於內視鏡特徵擷取不精確且隨機性大問題中，選用適合於內視鏡中之影像，並透過模型的學習及調整參數，使模型更適應於內視鏡影像當中，並結合色彩、亮度因素將內視鏡影像轉為不同色彩空間，使內視鏡影像能有物體遠近效果，於腫塊特徵上視覺效果上能夠更凸顯，讓模型學習該色彩空間特徵，內視鏡影像轉換為不同亮度，讓模型學習找出適合於內視鏡亮度環境之特徵，最後於僅參考單一辨識模型忽略其他重要分類模型中，改善僅參考單一辨識模型，加入本研究設計之深度學習 CNN、YOLO 辨識模型，並透過不同色彩影像之 CNN 集成、不同亮度之 CNN 集成、YOLO 集成及 YOLO 與 CNN 之多種排列組集成方式解決單一辨識模型之缺點，最後

於集成學習之結合多種分類器模型中進行優化輸出結果，使模型不以單一模型輸出結果作為依據，而是將多種模型之輸出作為參考，透過投票法則的方式得到最佳的分類結果。

經由以上改善，本研究期望輔助醫師能降低診斷鼻咽內視鏡影像長時間的問題，進而減少醫師的疲勞度，在降低醫師的疲勞的同時，也能讓病患有更好的診斷品質，在未檢驗出疾病的機率及風險都能夠有效的降低，進而提升醫療之品質。

本研究針對鼻咽內視鏡腫塊進行探討，研究範圍為鼻咽內視鏡含腫塊及不含腫塊之影像，鼻咽內視鏡影像則是使用 S 醫院部醫師所提供之鼻咽內視鏡影像資料集。

1.3 研究流程

本研究之流程如圖 2 所示，各階段說明如下：

1. 確認研究問題

藉由探討內視鏡辨識之研究議題，針對各研究辨識分類不足之處，並針對鼻咽腫塊探討辨識之可行性。

2. 相關文獻回顧

對於內視鏡之相關研究背景及發展進行回顧，並針對不同影像辨識之模型研究進行介紹，如傳統內視鏡影像辨識、現代內視鏡影像辨識。

3. 觀察資料

首先針對影像資料進行觀察，此步驟之目的為瞭解影像之像素、是否有雜訊、如何取感興趣區(Region Of Interest, ROI)使得模型有更好的效果及如何將腫塊特徵更為明顯。

4. 影像資料前處理

將影像經過裁切取得 ROI，影像中具強光雜訊之影像進行強光雜訊去除，避免模型學習強光之特徵，並將影像轉為不同色彩空間及不同亮度之影像以利突顯腫塊之特徵及擷取不同之特徵，將經影像前處理之影像進行標籤以輸入至模型訓練。

5. 開發影像辨識模型

建立影像辨識模型，最後透過集成學習對影像辨識模型進行學習，透過影像辨識模型最後預測結果利用集成學習投票方法得知該影像是否具有鼻咽腫塊。

6. 結論

根據實測之影像辨識模型之輸出結果來驗證本研究發展之影像辨識模型是否準確，並針對研究結果所提出具體之結論及未來研究建議。

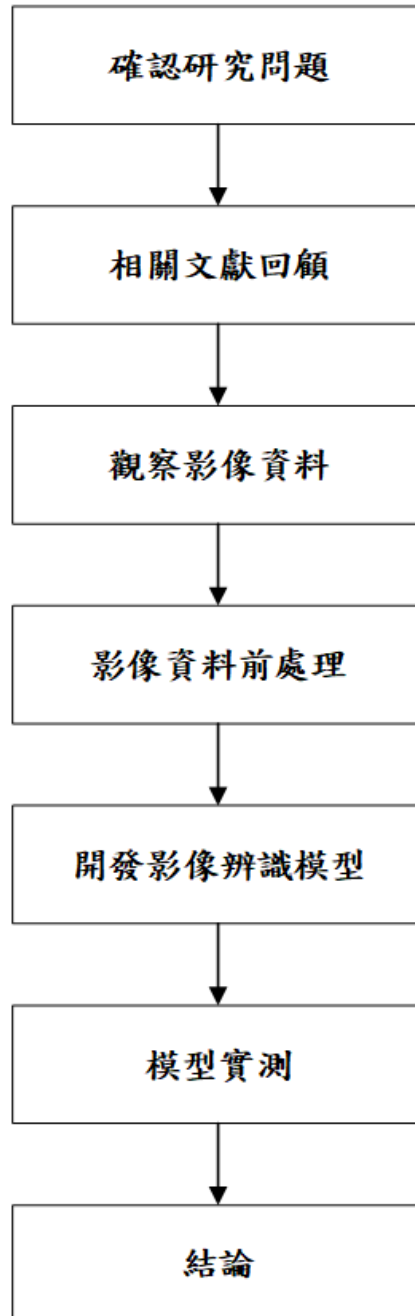


圖 2 研究流程

第二章 文獻回顧

2.1 基於機器學習影像處理及辨識方法

2.1.1 機器學習影像處理流程

基於機器學習辨識的流程為先利用特徵描述子演算法擷取影像之重要特徵，藉由特徵描述子演算法所索取之特徵轉換成直方圖，根據直方圖之特徵利用分類器對影像進行分類。常用於特徵描述子演算法有局部二值模式(Local Binary Patterns, LBP)、方向梯度直方圖(Histogram of Oriented Gradients, HOG)、尺度不變特徵轉換(Scale-Invariant Feature Transform, SIFT)，而常用於分類之方法有支援向量機(Support Vector Machine, SVM)及 K-Nearest Neighbors (KNN)，Yu 等人[62]於 2020 年研究關於子宮頸癌影像之辨識，期望在早期以低成本檢測子宮頸癌，並期望透過高精度的自動化檢測方法在病發前期提早檢測以便於有效預防子宮頸癌，該學者從 679 名患者中獲取的 4753 張陰道鏡影像的資料集，並透過傳統及基於深度學習影像處理及辨識方法針對其資料集進行訓練，最後以傳統的辨識方法準確度可以維持在 80%以上，敏感度則可以維持在 77%以上。Yuan 學者與 Meng 學者[65]於 2016 年，利用無線膠囊內視鏡進行息肉影像辨識，學者們使用傳統影像處理及辨識方法對影像進行處理，在特徵擷取階段，透過 SIFT 獲得息肉之特徵，並應用 K-means 進行分群，最後以自行研發之分類器進行分類，實驗結果能有效的辨識無線膠囊內視鏡之影像。以下小節將統整 LBP、HOG、SIFT、SVM、KNN 等方法，機器學習影像辨識流程如圖 3 所示。

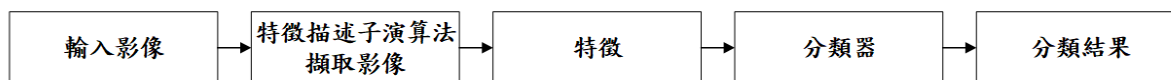


圖 3 基於機器學習影像處理及辨識流程圖

2.1.2 LBP

LBP 為區域性紋理特徵描述之方法，由 Ojala 等人[43]所提出，LBP 之作法如圖 4 所示，首先輸入一 3×3 大小之影像，以其影像中之中心為中心點，對周圍八點當作參

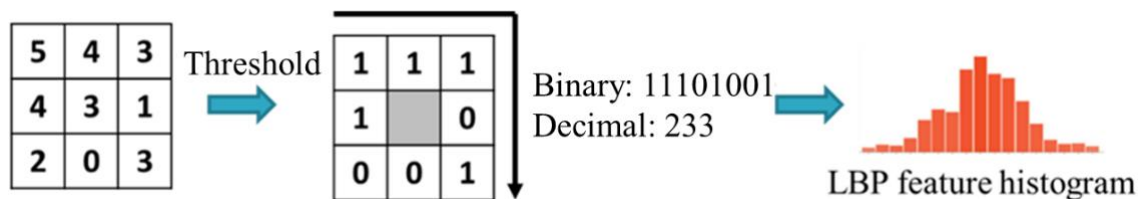


圖 4 LBP 流程圖

考點，接著比較參考點和中心點之像素值，並藉由中心點之像素值作為閾值對參考點去做二進制編碼，編碼的值可能為二元的 0(中心點大於參考點像素值)或者 1(中心點小於或等於參考點像素值)。完成編碼後，可以得到一個環狀二進制編碼，接著再以固定的起點及方向(逆時針或順時針)將環狀二進制編碼攤平成一個 8 位元的樣式，將 8 位元的樣式再轉為十進制之值，該值即為該點之特徵；若要敘述一個影像區塊，將計算影像區塊內所有點的特徵值，統計每個特徵值出現次數以直方圖的方式呈現，該直方圖即為影像區塊之特徵。以下為應用 LBP 擷取醫學影像特徵之研究，Kaplan 等人[30]於 2020 年應用 LBP 擷取腦瘤影像特徵，以 KNN、RF、類神經網路、線性判別分析 (Linear discriminant analysis, LDA) 當作分類器辨識其影像；Yu 等人[62]於 2020 年使用 LBP、HOG、SIFT 等特徵描述子擷取子宮頸癌影像特徵，以 SVM 當作分類器來辨識其影像；Korkmaz 與 Binol[34]於 2018 年使用 LBP、HOG 擷取胃癌影像特徵值，以隨機森林(Random Forest, RF)當作分類器來辨識其影像；Adjed 等人[4]於 2016 年使用 LBP 擷取皮膚癌影像特徵，以 SVM 當作分類器辨識其陰、陽性；De Sampaio 等人[18]於 2015 年使用 LBP 擷取 X 光之乳癌影像特徵，以 SVM 當作分類器來辨識其影像。

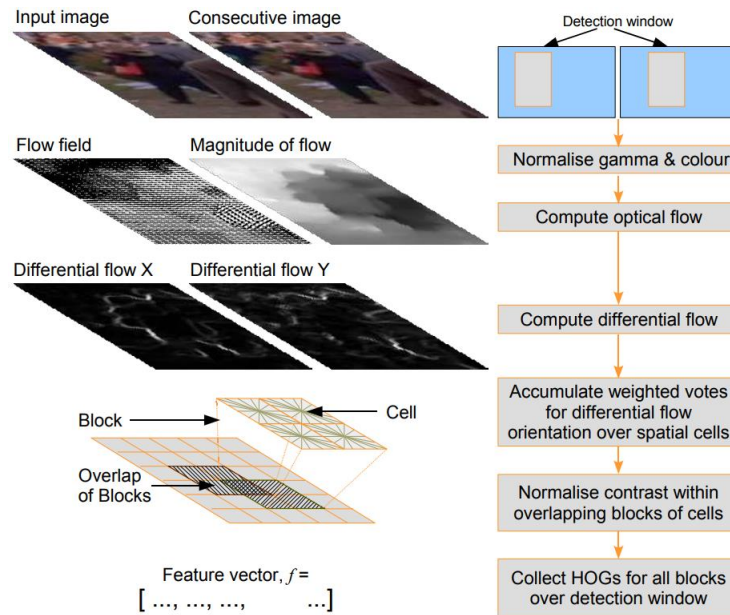


圖 5 HOG 流程圖[15]

2.1.3 HOG

HOG 由 Dalal 學者[15]於 2005 年提出，HOG 將影像轉為灰階以利於影像特徵擷取，採用 Gamma 校正法對輸入之影像值進行正規化，其目的是調節影像中的對比度，以利於降低影像局部之陰影及光照變化所進而造成的影響，採用 Gamma 校正法的同時亦可以降低雜訊的干擾，計算影像每一個像素的梯度(即為對每一個像素計算斜率)，透過計算梯度的方式可以得知該像素之大小和方向，其主要目的是為了捕捉輪廓特徵，進一步弱化光照的干擾，對影像劃分成 $G \times G$ 個網格(如 6×6 之網格)，統計每一個網格的梯度直方圖，即可形成每一個網格的特徵描述。HOG 流程圖如圖 5 所示。以下為應用 HOG 擷取醫學影像特徵之研究，Savitha 與 Jidesh[52] 於 2019 年使用 HOG 擷取 CT 之肺結核影像特徵並利用 K-means 將特徵分群，以 SVM 當作分類器分類其陰、陽性；Alfed 與 Khelifi[6]於 2017 年使用 HOG、紋理、色彩特徵擷取皮膚癌影像特徵，以 SVM、類神經網路當作分類器辨識其陰、陽性；Bakheet[11]於 2017 年使用 HOG 擷取皮膚黑色素癌影像特徵，以 SVM 當作分類器分類其陰、陽性；Iwahori 等人[28]於 2015 年使用 HOG 擷取內視鏡之大腸息肉影像特徵，以 RF 當作分類器分類其陰、陽性；David 等人[17]於 2013 年使用 HOG 擷取內視鏡之大腸息肉影像特徵，以類神經網路當作分類器分類其陰、陽性。

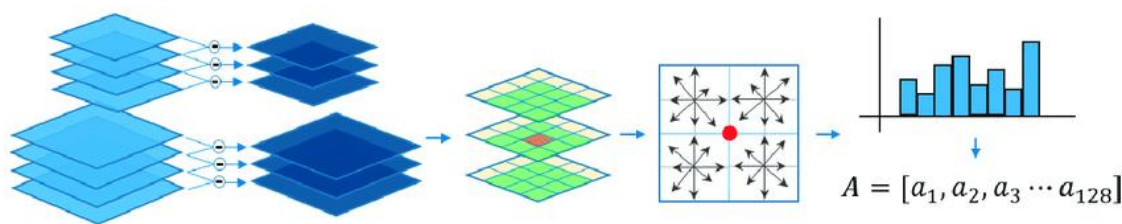


圖 6 SIFT 流程圖[40]

2.1.4 SIFT

SIFT 由 Lowe[40]於 1999 年提出，主要在不同的尺度空間中找尋關鍵點即特徵點，並計算出該關鍵點之梯度方向，而 SIFT 的擷取特徵主要過程是用了高斯差(Difference of Gaussians, DoG)，即將影像轉為不同尺度的高斯模糊，藉由影像的縮放取的紋理特徵中比較複雜的區塊(例如：邊緣、點、角等等)，接著透過如 HOG 的方法，對每一個網格之像素值取梯度，即計算其梯度方向及大小，統計每個梯度方向特徵值出現次數以直方圖的方式呈現，該直方圖即為影像區塊之特徵，SIFT 流程圖如圖 6 所示。以下為應用 SIFT 擷取醫學影像特徵之研究，Gheisari 等人[22]於 2018 年使用 LBP、SIFT 擷取神經母細胞瘤影像特徵，以 SVM 當作分類器分類其陰、陽性；Shiji 等人[54]於 2017 年使用 SIFT 及 BoF 擷取超音波之乳癌影像特徵，以 SVM 當作分類器分類其陰、陽性；Yuan 與 Meng 學者[65]於 2016 年將 SIFT 應用於 HSV 色彩空間之內視鏡之大腸息肉影像擷取特徵，以 SVM 當作分類器分類其陰、陽性；Yuan 於 2015 年等人[64]使用 LBP、HOG、SIFT 擷取內視鏡之大腸息肉影像，以 SVM 當作分類器分類其陰、陽性；Désir 等人[20]於 2012 年使用 LBP、SIFT 擷取顯微肺結核影像特徵，以 SVM 當作分類器分類其陰、陽性。

2.1.5 SVM

SVM 為一種監督式的學習方法，在線性分類及非線性分類都可以很有效的運用，其主要目的在於找尋一條線，使類別與類別之間之距離能夠越分離，同時也找尋的最大邊界距離，如圖 7 給定一組由藍色正方形及紅色三角形之訓練資料集，正方形及三角形個表示一個類別，以 SVM 之概念去訓練建立一條線，可以看見正方形及三角形兩類別能透過該條線及邊界盡可能分開，若有新的資料要輸入至 SVM 中，即會根據該形成之間隔來預測所屬之類別，其主要公式如式(1)所示。以下為應用 SVM 之醫學影像辨識研

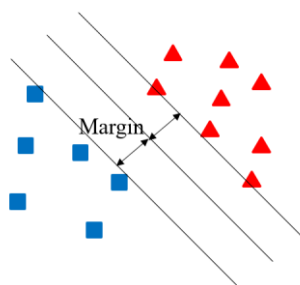


圖 7 SVM 示意圖

究，Khan 等人[32]於 2020 年使用紋理、點和幾何擷取 CT 之肺癌影像特徵，以 SVM 當作分類器分類其陰、陽性；Vijayarajeswari 等人[61]於 2019 年使用霍夫變換(Hough transform)擷取 X 光之乳癌影像特徵，以 SVM 當作分類器分類其陰、陽性；Amin 等人[8]於 2019 年使用小波轉換、LBP 擷取 MRI 之腦瘤影像特徵，以 SVM 當作分類器分類其陰、陽性；Deepika 等人[19]於 2019 使用 LBP 擷取 MRI 之腦瘤影像特徵，並使用主成分分析(Principal components analysis, PCA)進行降維，最後以 SVM 當作分類器分類其陰、陽性；Li 等人[37]於 2012 年使用 LBP 及小波轉換擷取內視鏡之大腸息肉影像，以 SVM 當作分類器分類其陰、陽性。

$$f(x) = w^T x + b \quad (1)$$

2.1.6 KNN

KNN 是一種分類的方法，即輸入一個訓練資料集，在訓練資料集中找到與該輸入資料點中最鄰近的 K 個點，如果這 K 個點多數屬於某一個類別，意即將新輸入之資料點分類到此類別當中，如圖 8 所示，有兩個不同類別的樣本資料，分別使用藍色的小正方形及紅色的小三角形表示，而圖中的綠色的圓所代表的欲分類之輸入資料點，根據 K-近鄰演算法的少數服從多數的想法來給綠色圓點進行分類，若 $K=3$ ，綠色圓點的最鄰近的 3 個點中有 2 個紅色小三角形及 1 個藍色小正方形，少數服從多數，即將綠色點分類至紅色三角型之類別；若 $K=5$ ，綠色圓點的最鄰近的 5 個鄰近點是 2 個紅色三角形及 3 個藍色的正方形，少數從屬於多數，即將綠色點分類至藍色正方形之類別，其公式如式(2)所示。以下為應用 KNN 之醫學影像辨識研究，Mishra 等人[41]於 2019 使用 SIFT 及 SURF 擷取 X 光之乳癌影像特徵，並透過 PCA 進行降維，以 KNN 作為分類器

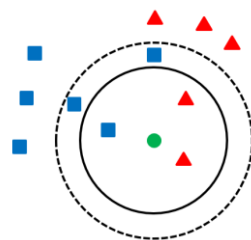


圖 8 KNN 示意圖

分類其陰、陽性；Kavitha 等人[31]於 2017 年使用灰度共生矩陣(Gray Level Co-occurrence Matrix, GLCM)、加速穩健特徵(Speeded up robust Features, SURF)擷取皮膚黑色素瘤影像，以 SVM 和 KNN 當作分類器分類其陰、陽性；Amulya 與 Prathibha[10]於 2016 年使用 SURF、SIFT 擷取 MRI 之腦瘤影像特徵，以 KNN 作為分類器分類其陰、陽性。

$$Sim(d_i, d_j) = \frac{\sum_{k=1}^M W_{ik} \times W_{jk}}{\sqrt{(\sum_{k=1}^M W_{ik}^2)(\sum_{k=1}^M W_{jk}^2)}} \quad (2)$$

表 1 基於機器學習影像處理及辨識相關文獻

作者(年份)[參考文獻]	目標	方法
Yu 等人(2020)[62]	子宮頸癌影像	SIFT、SURF、ORB、LBP、 HOG、SVM、RF、KNN
Korkmaz 與 Binol(2018)[34]	胃癌影像	LBP、HOG、RF
Prasath 學者(2017) [46]	內視鏡之大腸息肉影像	SIFT、SURF、ORB、LBP、 HOG、SVM、RF、KNN
Cheng 等人(2017) [14]	MRI 腦瘤影像	BoF、KNN、SVM
Yuan 等人(2016)[65]	內視鏡之大腸息肉影像	SIFT、LBP、BoF、SVM
Yuan 等人(2015)[64]	內視鏡之大腸息肉影像	SIFT、SVM
Kumar 等人(2015)[35]	微觀影像	HOG、KNN、SVM
Gueye 等人(2015)[24]	內視鏡之大腸息肉影像	SIFT、BoF、K-means、SVM

2.2 基於深度學習影像處理方法介紹

2.2.1 深度學習影像處理流程

相較於應用描述子影像辨識的處理流程，深度學習影像辨識不必再透過描述子先對影像進行特徵擷取的動作，而是直接利用深度學習的方式學習輸入影像之特徵，深度學習中會透過卷積層進行特徵擷取的動作，其動作就如描述子對影像特徵擷取的方式，接著透過類神經網路的學習，學習該卷積層所擷取之特徵，最後藉由學習之特徵網路預測出該影像之類別，圖 9 為基於深度學習影像辨識之流程圖。Safdar 等人[50] 於 2020 年對腦部進行腦瘤 MRI 影像辨識，該研究收集了 1961 張影像，並透過資料增強的方式增加影像樣本，最後透過基於深度學習方法 YOLOv3 進行辨識，最後該實驗能有效的辨識該疾病。Lee 等人[36]於 2020 年對大腸內視鏡影像辨識是否有息肉的部分，該學者們透過端對端之 CNN 模型進行模型上的微調，最後將影像直接輸入至 CNN 模型進行辨識，最後證實其結果是有效的。Yu 等人[62]於 2020 年研究關於子宮頸癌影像之辨識，從 679 位病人收集子宮頸癌影像，最後收集 4753 張陰道鏡影像的資料集，並藉由傳統及基於深度學習影像處理及辨識方法針對其資料集進行訓練，在深度學習的部分使用了常見之 CNN 架構，實驗結果發現傳統影像處理及辨識方法遠低於基於深度學習之方法。

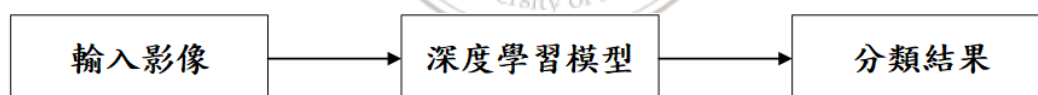


圖 9 深度學習之影像辨識流程圖

2.2.2 卷積類神經網路(Convolutional Neural Network, CNN)

在神經網路當中，卷積類神經網絡是進行影像辨識、影像分類及物件偵測的主要方法之一，其中如臉部、醫療影像辨識等等，在 CNN 中皆廣泛運用於這些領域。CNN 的影像分類方法藉由已標籤之輸入影像，將輸入影像視為像素的陣列，從技術上來說，深度學習 CNN 模型應用於訓練及測試當中，每個輸入之影像將通過一系列帶有卷積層中之卷積核(Kernels)，池化層(Pooling)，全連接層(Fully Connected, FC)的網路傳遞，並透過 Softmax 函數對欲分類之影像給予類別機率值，透過類別機率值賦予分類之標籤，下圖 10 為 CNN 處理輸入影像並基於影像中的陣列值對影像進行分類之示意圖。以下為應

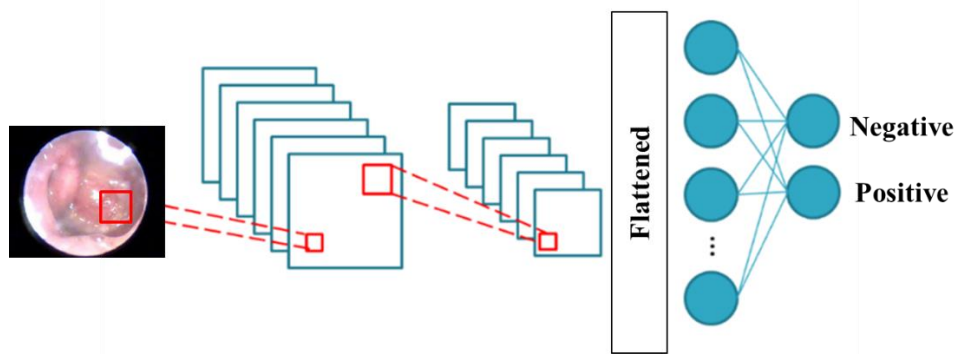


圖 10 CNN 之示意圖

用 CNN 架構之醫學影像辨識研究，Yu 等人[62]於 2020 年使用 AlexNet、VGG16、GoogleNet、InceptionV3、ResNet 辨識子宮頸癌影像之陰、陽性；Urban 等人[60]於 2018 年使用 CNN 辨識內視鏡之大腸息肉影像陰、陽性；Bardou 等人[12]於 2018 年使用 CNN 辨識 X 光之乳癌影像陰、陽性；Tajbakhsh 等人[58]於 2017 年使用微調之 CNN 模型辨識內視鏡之大腸息肉影像陰、陽性；Amit 等人[9]於 2017 年使用 CNN 辨識 MRI 之乳癌影像陰、陽性；Halicek 等人[26]於 2017 年使用 CNN 辨識腦部高光譜影像是否具有腦癌特徵。

2.2.3 YOLO

You Only Look Once (YOLO)模型為目前物件偵測當中運算速度最快之一的模型，同時也成為現今常用的物件偵測模型，自 2016 至 2018 推出了三個版本，由 Redmon 等人[49]於 2016 年提出，YOLO 將物件偵測視為一個單一迴歸問題，YOLO 先將輸入的圖像劃分成 $G \times G$ 的網格，接著透過基於迴歸之 CNN 模型為每個網格產生出邊界框 (Bounding Box)，如果網格中包含著欲分類之物件，則會為每個網格預測出物件於該網格出現的機率值以及該物件類別機率值，最後透過非極大抑制 (Non-Maximum Suppression, NMS) 去刪取多餘的邊界框，剩下來的邊界框即為最後的預測框，而預測框將會與物件的真實框藉由損失函數進行模型的修正來取得更精確的位置之物件。

由 Redmon 與 Farhadi [47]於 2017 年提出 YOLO 第二個版本，以 YOLOv1 作為基礎，在每個卷積網路層後添加了批次正規化，讓權重值能夠不偏向於其中任一項權重值，並且加入了避免過度擬和的 Dropout，在輸入尺寸從原本 7×7 更改為 13×13 ，其目的在

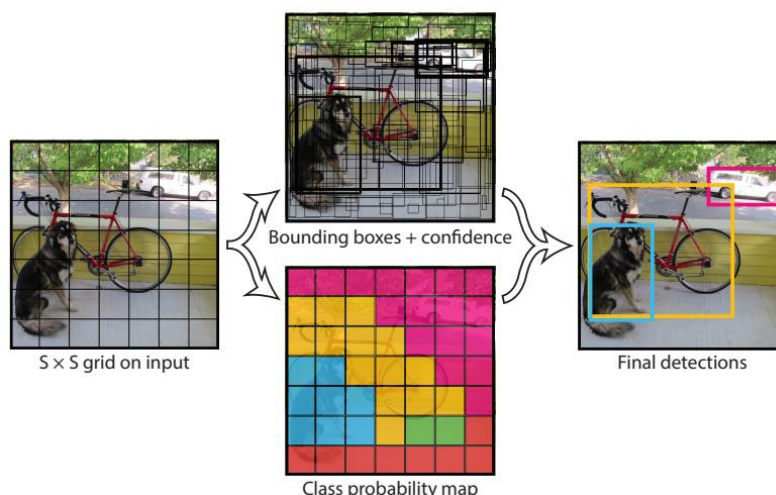


圖 11 YOLO 模型概念圖[49]

於提高尺寸能學習到影像中細節的部分，而為了克服 YOLOv1 每一個網格僅能預測一個物件類別的缺陷，借鑑 Faster R-CNN 中錨點框(Anchor Box)的思路，放棄了原本使用全連接層來預測之預測框，而是藉由錨點框的偏移量(Offset)及 IoU 來獲得最後的預測框，同時每個錨點框都為預測每一個類別，而在 Faster R-CNN 錨點框的長寬值都是由人工設定的，YOLOv2 則是利用 K-Means 的方式對影像資料集之邊界框做分群來獲得長寬值。

由 Redmon 與 Farhadi [48] 於 2018 年提出 YOLO 第三個版本，YOLOv3 是採用了 YOLOv2 作為基礎，從網路的骨架及輸出層的結構上做了改良，使得 YOLOv3 在準確度上有著大幅的提升，其主要改良的思路是有更多輸出尺寸的預測框，其輸出之預測框尺寸改為 13×13 , 26×26 以及 52×52 三種，尺寸越大代表能劃分出更細緻的網格，就能從影像中檢測出越精細的物件，加深了網路骨架，提出了新的 Darknet-53 骨架，53 即代表 53 個卷積層，採用簡單的殘差區塊作為加深網路的方法代替 YOLOv2 中最大池化層，並且結合了特徵金字塔網路 (Feature Pyramid Networks, FPN)[39]的概念，利用從下而上(bottom-up)、從上而下(up-bottom)以及橫向連結(lateral connections)之路徑提取影像中特徵值，藉由 FPN 多層級預測的架構以有效的提升小物體中細節預測能力，使用 FPN 的架構可以讓低層較佳的目標位置和高層較佳的特徵融合，並且在不同特徵層獨立進行預測，使得小物體檢測改善效果十分明顯，而在 YOLO 系列中，僅有 YOLOv1 有明確的

提出損失函數之公式，而 YOLOv3 有將原本 YOLOv2 的損失函數 Softmax 改成 Logistic 損失函數，其主要的目的在於 Softmax 改良只能對應至一個類別的問題，圖 11 為 YOLO 概念圖。以下為應用 YOLO 架構之醫學影像辨識研究，Safdar 等人[50]於 2020 年對 MRI 腦部腫瘤影像進行辨識；Cao 等人[13]於 2019 年對超音波乳癌影像進行辨識；Al-Masni 等人[7]於 2018 年對 X 光之乳癌影像進行辨識；George 等人[21]於 2018 年對 CT 之肺結核影像進行辨識；Pogorelov 等人[45]於 2017 年使用 YOLO 對內視鏡之大腸息肉影像進行辨識。

非極大抑制(Non-Maximum Suppression, NMS):

在物件偵測的方法中時常會發生出現多個邊界框的問題，NMS 為物件偵測篩選邊界框常用的方法之一，透過 NMS 來篩選出最合適物件類別之邊界框，NMS 會根據從網格所預測多個之邊界框挑選出對物件有信心程度最高之邊界框，其他邊界框則會從剛選出來之信心程度最高的邊界框來計算 Intersection over Union(IoU)，接著依據 IoU 是否大於預先設定好的閾值之邊界框來為其他邊界框計算信心分數，如果算出來之 IoU 大於設定好之閾值，則將 IoU 大於閾值之候選框的信心分數設定為 0，一直做到每個類別都有挑選出其最高之邊界框為止，其公式如式(3)所示。

$$s_i = \begin{cases} s_i, & IoU(M, b_i) < N_i \\ 0, & IoU(M, b_i) \geq N_i \end{cases} \quad (3)$$

Intersection over Union (IoU):

IoU 為一種測量在資料集中檢測物體準確度的一個指標，其公式如式(4)所示，IoU 定義為：

$$IoU = \frac{ground\ truth \cap predict\ result}{ground\ truth \cup predict\ result} \quad (4)$$

Ground-Truth 為真實框之面積，Predict Reult 為預測框之面積，圖 12 為 IoU 之視覺化圖像示意圖。

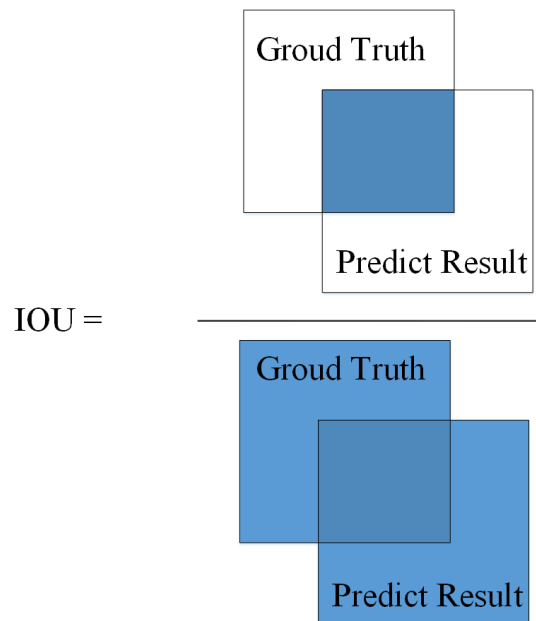


圖 12 IoU 視覺化示意圖

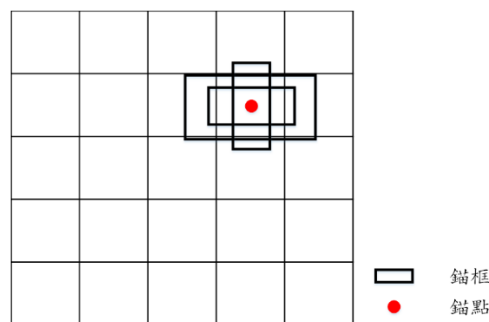


圖 13 錨點示意圖

錨點(Anchors)

由於過往 YOLOv1 每一個網格只能預測一個物件，若影像中有物件重疊的情況發生，會導致模型在預測上的困難，為了解決上述的困難，將物件之中心點設置為錨點，並設定尺寸不同之錨框(Anchor Box)，錨框及錨點如圖 13 所示。

K-Means

K-means 演算法，為一種非監督式分群分析之演算法，目標為將下列公式(5)達到最小化，即將資料分為 k 個群組，即計算每個資料點分配到的群重心點距離總合為最小。依據 K-means 的想法，首先隨機選擇資料集中 k 個點作為初始群重心，即 μ_1 到 μ_k ，透過

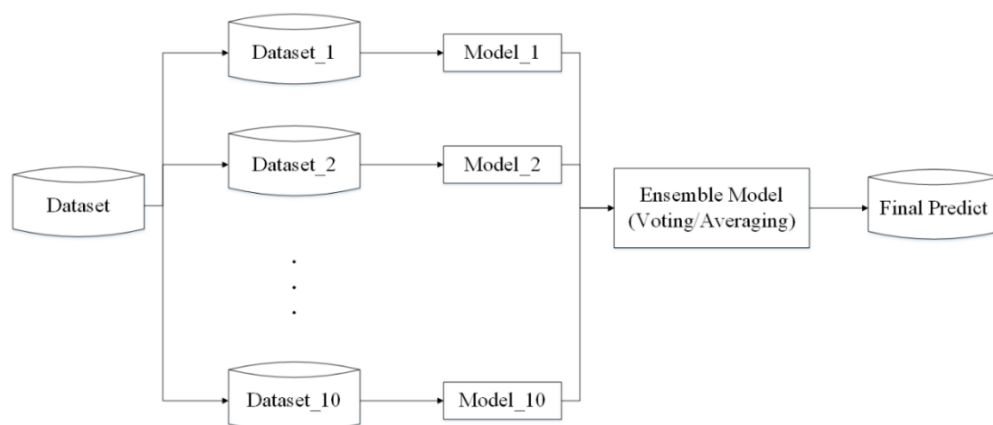


圖 14 Bagging 範例圖

群重心與每個資料點 x_i 計算最短距離的，計算完畢後更新群中心，藉由目前所分群之集群再次以群重心計算，不斷的重複上述步驟直到達到收斂為止。

$$J = \sum_{i=1}^k \sum_{x_j \in S_i} (x_j - \mu_i)^2 \quad (5)$$

表 2 基於深度學習之醫學影像辨識之相關文獻

作者(年份)[參考文獻]	辨識項目	方法
Yu 等人(2020)[62]	子宮頸癌影像	AlexNet、VGG16、GoogleNet、InceptionV3、ResNet
Pogorelov 等人(2018)[44]	內視鏡之大腸息肉影像	Xception, VGG19, ResNet50
Jha 等人(2020)[29]	內視鏡之大腸息肉影像	YOLO
Lee 等人(2020)[36]	內視鏡之大腸息肉影像	CNN
Safdar 等人(2020)[50]	MRI 腦部腫瘤影像	YOLO
Akbari 等人(2018)[5]	內視鏡之大腸息肉影像	CNN

2.3 集成式學習在影像中的處理方式

集成學習為以一種系統化的方式以多種監督式學習結合而成的一種模型，其目的是為了從多個模型之結果進行整合，透過整合的模式去影像單一模型之準確度或方差的結果，以達到穩定的提升最終模型之績效的一種方式。

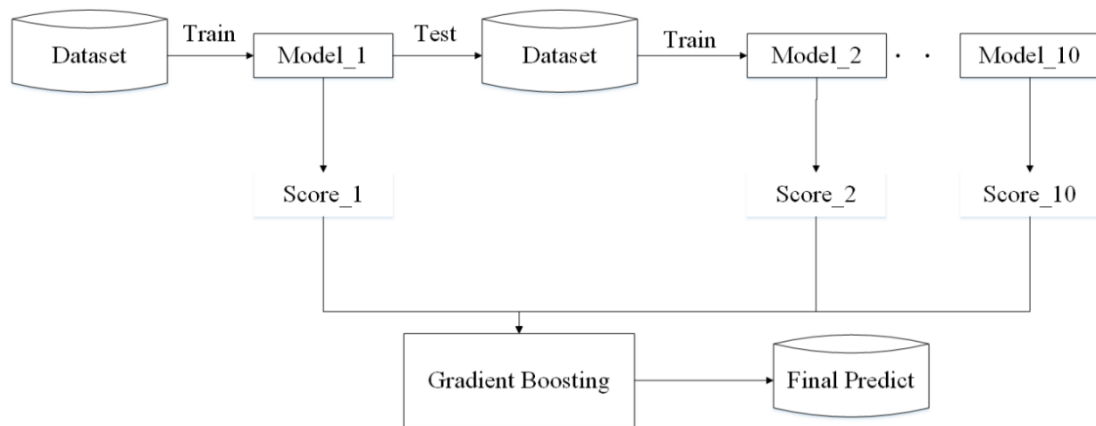


圖 15 Boosting 範例圖

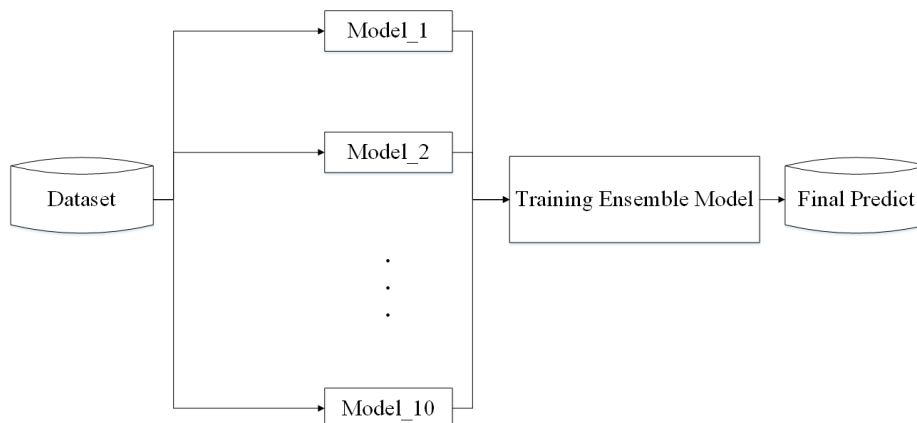


圖 16 Stacking 範例圖

常見的集成學習方法分別有 Bagging、Boosting、Stacking 以上三種，Bagging 為透過將資料集進行不同採樣而產生之訓練集對多個模型進行訓練，從多個模型當中之結果以投票法之方式來獲取最終結果，圖 14 為 Bagging 之範例；Boosting 為不將資料集重新採樣，而是以相同訓練集進行訓練，其目的是為了避免單一模型在訓練當中有些資料在判斷上永遠都是錯的，若又將所有的資料再進行訓練，錯誤的資料在判斷上永遠都是判定成錯誤的，因此 Boosting 是針對錯誤的資料去訓練，從訓練的過程中將錯誤的資料權重加大，那麼新訓練出來的模型才能從錯誤資料得到一個較佳的結果，圖 15 為 Boosting 之範例；Stacking 則為將訓練多個不同模型進行整合，透過各模型輸出之結果再去訓練一個模型得到最終之結果，圖 16 為 Stacking 之範例。Liew 等人[38]於 2021 年透過 Boosting 辨識大腸內視鏡影像之息肉陰、陽性；Yu 等人[62] 於 2020 年經由 Boosting 的方式辨識子宮頸癌影像之陰、陽性。

表 3 基於集成學習之醫學影像辨識之相關文獻

作者(年份)[參考文獻]	辨識項目	方法
Liew 等人(2021)[38]	內視鏡之大腸息肉影像	Boosting
Ghosh 等人(2021)[23]	內視鏡之大腸息肉影像	Stacking
Guo 等人(2019)[25]	內視鏡之大腸息肉影像	Stacking
Yu 等人(2020)[62]	子宮頸癌影像	Boosting

2.4 小結

本研究將第 2 章模型的部份分為基於傳統及基於深度學習影像辨識方法，2.1 小節統整基於機器學習之特徵描述子於影像之辨識演算法，2.2 小節介紹深度學習內視鏡影像之辨識演算法，2.3 小節介紹集成學習相關之影像辨識演算法。從過往研究顯示，基於統計方法及傳統之特徵描述子影像辨識的效果並沒有透過深度學習影像辨識的成效要好，其原因為描述子特徵擷取的方式在深度學習中之卷積層中都可以做到，且再透過全連接層網路的學習，使得模型更能學習出卷積層所擷取之特徵，在效果上深度學習影像辨識之方法明顯大過於特徵描述子影像辨識，本論文後續將以 2.2 及 2.3 小節中介紹深度學習之內視鏡影像辨識之 CNN 及 YOLO 作為主要使用模型，透過 CNN 及 YOLO 最後輸出之結果進行集成學習，並修改其中之參數與模型架構進行模型訓練。

第三章 資料集與初步整理

本章節介紹由 S 醫院醫師提供之鼻咽內視鏡影像資料集，其原始資料為紙本資料，需將其轉為電子資料集方式以供模型進行訓練，該影像資料集總張數為 234 張，其中陽性總筆數為 127 筆影像，陰性總筆數為 107 筆影像。

3.1 紙本資料

本研究所使用 S 醫院醫師所提供之鼻咽內視鏡資料集，其分為紙本與未標籤之影像資料集，紙本中若是醫師於影像中圈選患部位置則將該影像標籤為 1，即陽性的部分，相反的，若是無圈選患部位置則將該影像標籤為 0，即陰性的部分。整理之電子資料集於 3.2 節中介紹。

3.2 CNN 內視鏡鼻咽部位資料

整理之 CNN 內視鏡鼻咽影像資料集，其資料欄位名稱為影像名稱、影像類別(Class)0 為陰性、1 為陽性，如表 4 CNN 內視鏡鼻咽部位資料集所示，而陰性及陽性如圖 17、圖 18 所示。

表 4 CNN 內視鏡鼻咽部位資料集

影像名稱	Class
0475311_0810211420_I00064468876.jpg	0
0547606_0704090615_I00052501407.jpg	0
0208952_0603300639_I00045164832.jpg	0
0208952_0603300639_I00045164833.jpg	0
...	...
0045289_0510172120_I00041884860.jpg	1
0009911_0804032249_I00060161229.jpg	1
7961601_0512071474_I00042902874.jpg	1
7961601_0512071474_I00042902874.jpg	1

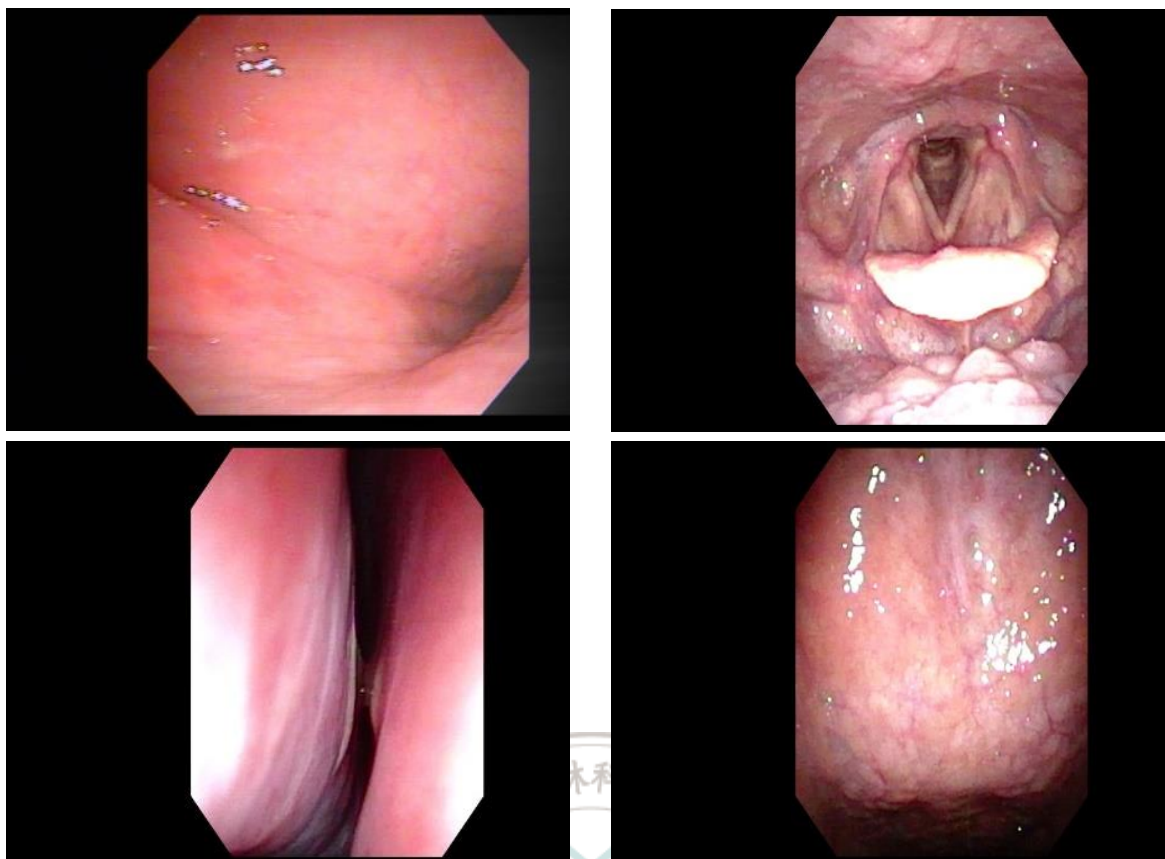


圖 17 鼻咽陰性範例圖

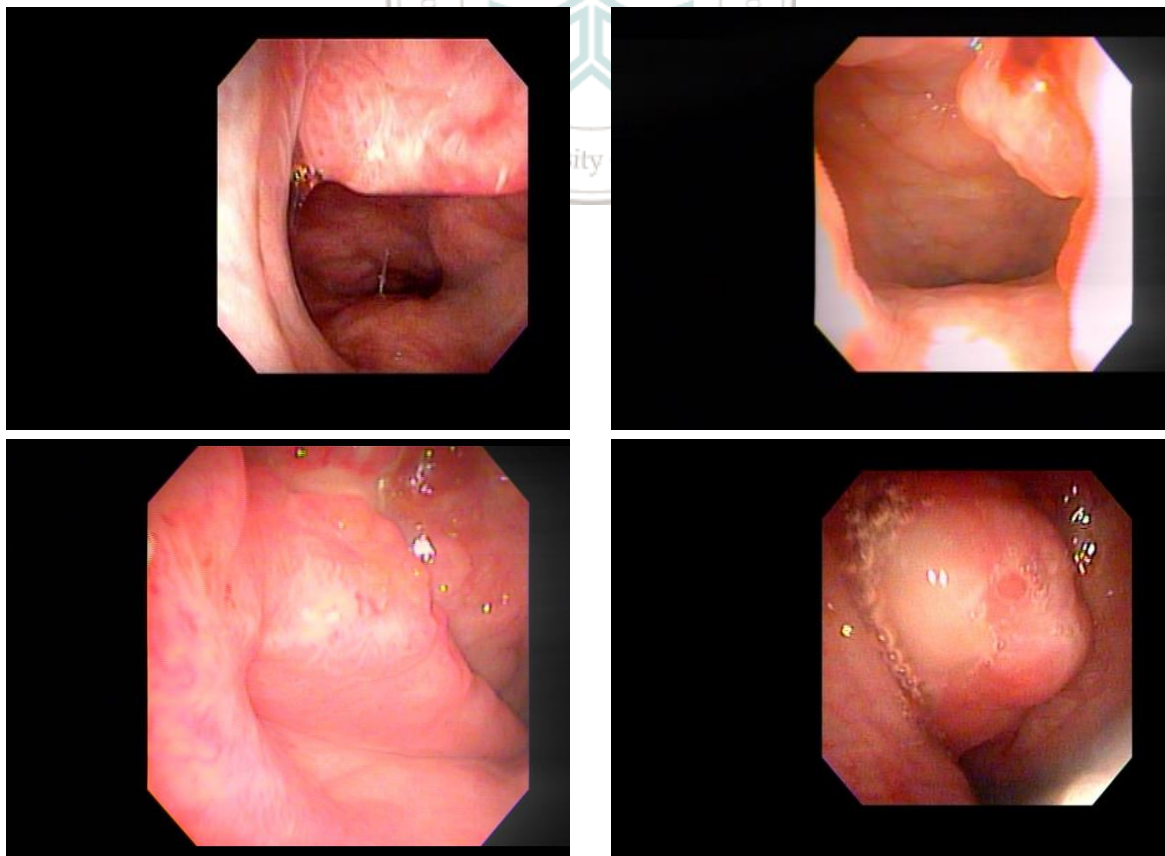


圖 18 鼻咽陽性範例圖

第四章 基於 CNN 模型之做法

本論文提出之模型概念為以 CNN 及 YOLOv3 兩種方法，針對原始色影像、不同色彩模型影像、不同亮度影像三種影像之特徵分別以 CNN 模型進行訓練，而 YOLOv3 則僅以原始色影像進行訓練，本論文將模型訓練分為 Offline 訓練及 Online 運作兩部分兩個部分，分別為如圖 19 所示，在 Offline 運作部分，第一階段為內視鏡影像前處理的階段，取出 ROI 影像區塊，並去除內視鏡影像中強光之雜訊，去除強光雜訊後，將影像轉為不同色彩空間及不同亮度之影像，最終有三種影像輸入至模型之資料集。第二階段為此三種資料集分別設計之 CNN 模型，各別 CNN 模型嘗試擷取各影像資料集之特徵。第三階段為以集成學習透過投票的方式去控制三種影像資料集之分類，使集成學習模型依據三種影像資料集 CNN 預測之特徵，透過投票的方式判別該影像的類別，因應各 CNN 預測之特徵變化做出最後的分類。因此本研究以第一階段至第三階段之訓練為主要重點，透過集成學習找出良好之分類類別，YOLOv3 的部分將會於第五章進行說明。在 Online 運作部分，則會依據 Offline 所挑選之特徵值進行應用於實際鼻咽內視鏡影像當中。

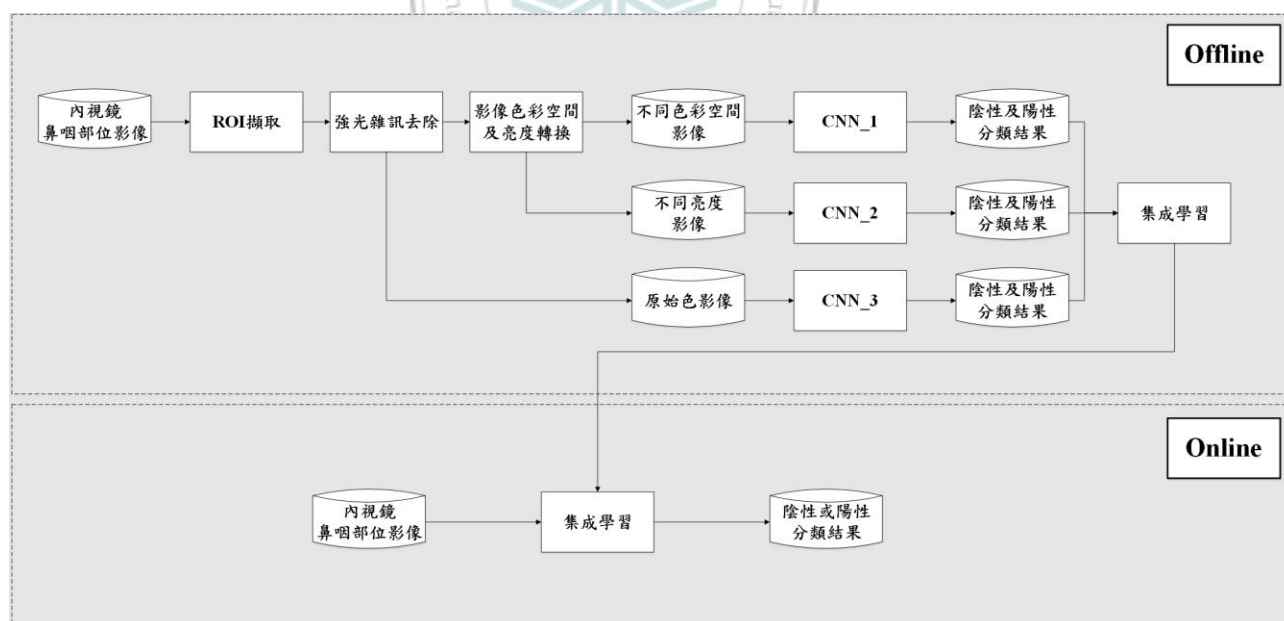


圖 19 CNN 研究流程圖



圖 20 取出 ROI 區塊流程圖

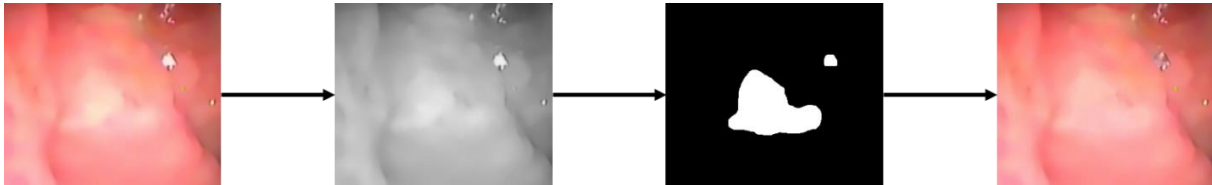


圖 21 強光雜訊去除流程圖

4.1 擷取 ROI

在我們所獲得的內視鏡影像資料集中皆有其獨特之黑色外框，為了使模型分類能力更加的準確，避免模型因黑色外框的因素而造成準確率下降，因此，先將影像轉為灰階以利抓取黑色外框部分，透過二值化分割出黑色外框及非黑色外框區塊，二值化方法為藉由閾值將閾值以上之像素值設定為 255，將閾值以下之像素值設定為 0，接著抓取非黑色外框之輪廓，並透過輪廓獲得其寬、高位置值，接著透過輪廓中寬、高位置值以影像裁切方式裁出 ROI 區塊並將尺寸大小設置為 300×360 ，擷取 ROI 區塊流程如圖 20。

4.2 強光雜訊去除

去除黑色外框後，還需要去除內視鏡於人體內所產生之強光反射雜訊，強光反射之雜訊也會影響模型在網路中權重之訓練，因此，將 ROI 區塊之影像轉為灰階以利抓取強光雜訊，並設定適應於抓取強光雜訊之二值化閾值，透過二值化的方式獲得強光遮罩，接著透過 Telea [59]補值方法對強光雜訊遮罩進行補值，該補值方法會取得強光雜訊遮罩周圍之影像像素，藉由周圍影像像素對強光雜訊遮罩白色處進行補值，直到白色處皆無為止，最後取得去除強光雜訊之影像，強光雜訊去除流程如圖 21 所示。

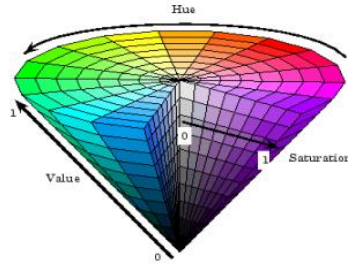


圖 22 HSV 圓柱坐標系

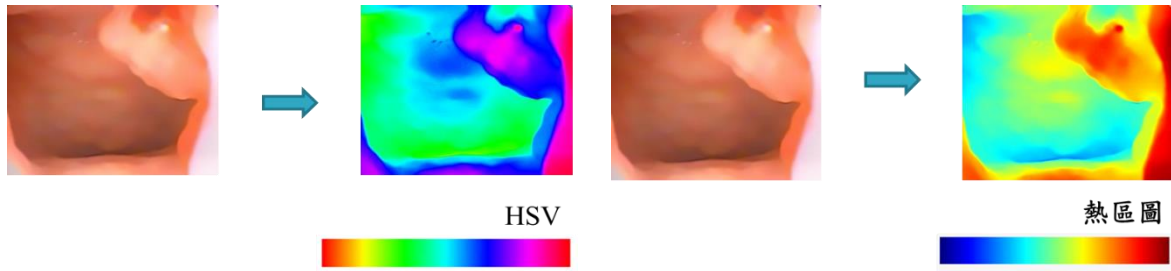


圖 23 原始色影像轉不同色彩空間示意圖



圖 24 原始色影像轉不同亮度示意圖

4.3 影像色彩空間及亮度轉換

強光雜訊去除後，藉由轉換不同色彩空間及不同亮度方式，使得模型能夠擷取更多不同特徵值，本論文將原始色影像轉換為 HSV 色彩空間、熱區圖色彩空間及不同亮度影像，HSV 色彩空間中字母 H 為色相、字母 S 為飽和度、字母 V 為明度，該色彩空間是由三原色(RGB)中的點在圓柱坐標系中的一種表示法，HSV 圓柱坐標系如圖 22 所示，透過 HSV 色彩空間能使影像在視覺中看起來有物體遠近視覺效果，而熱區圖色彩空間亦是如 HSV 相同之表示法，但是其色彩空間值較窄，亦是將影像在視覺中看起來有物體遠近視覺效果，圖 23 為轉變不同色彩空間示意圖，圖 24 為轉變不同亮度示意圖。



圖 25 資料增強示意圖

4.4 資料增強設計

由於 CNN 模型需要一定的樣本數才足以學習到該影像資料集特徵，且為了避免 CNN 模型在訓練時有過擬和的情形發生，本論文在樣本數較為不足的情況下，使用資料增強的方式來去彌補資料樣本數不足的狀況，而常見的資料增強的方式有將影像進行旋轉、翻轉、裁切等方式來增加樣本數及樣本的多樣性[55]，資料增強範例如圖 25 所示。

4.5 建立基礎影像辨識之 CNN 模型

4.5.1 CNN 集成學習之模型選擇

取得原始色影像、不同色彩空間影像及不同亮度影像後，各別影像會輸入至各別 CNN 模型中，透過 CNN 模型輸出之結果並藉由集成學習得到較好的結果，以下為集成學習後之流程，輸入為過往 CNN 訓練結果的類別特徵(意即 0 與 1)，輸入至集成學習模型當中，輸出則為根據集成學習中模型透過少數服從多數的投票法則做出最後的分類預測。若本論文設計多個 CNN 模型，將多個模型之預測分類輸入至集成學習模型當中，集成學習模型會依據各別 CNN 輸出之結果透過集成學習模型再各別預測分類值，最終集成學習模型依據投票法之法則輸出最後分類結果，本實驗將設計不同色彩空間、不同亮度及混和亮度及色彩空間之集成學習，如圖 26、27、28 所示。以下章節將詳細介紹 CNN 模型架構模型部分及集成學習框架的部分。

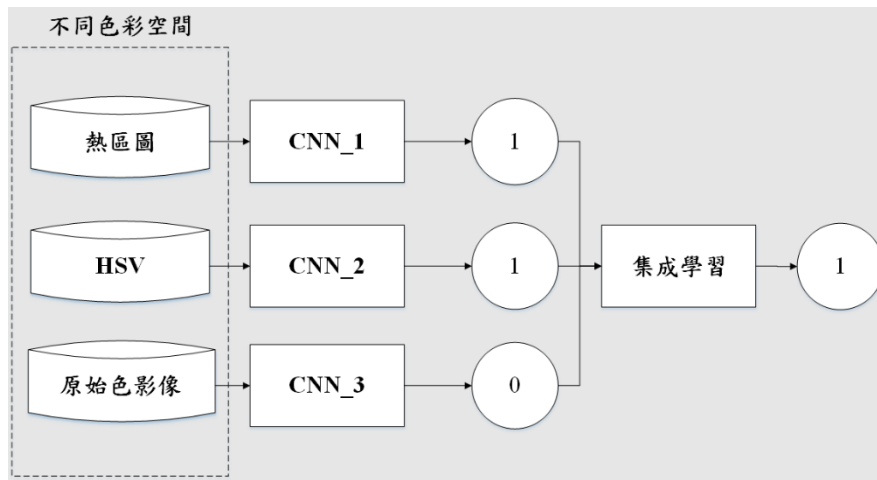


圖 26 CNN 不同色彩空間影像集成學習

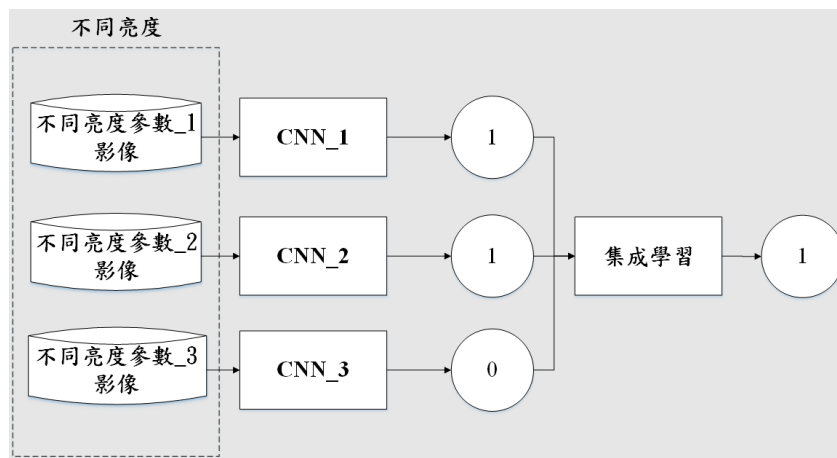


圖 27 CNN 不同亮度影像集成學習

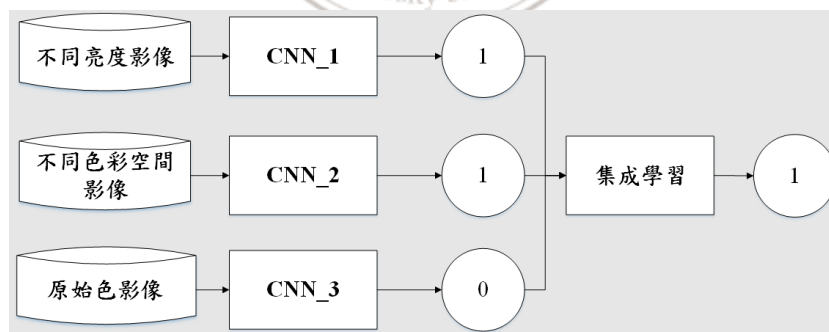


圖 28 CNN 混和式集成學習

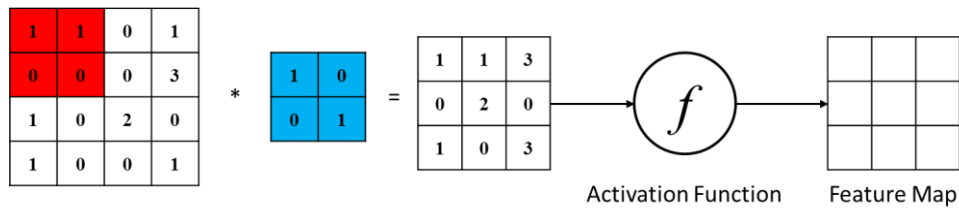


圖 29 卷積過程示意圖

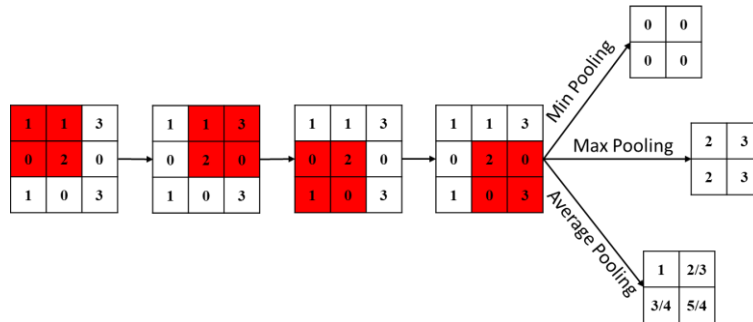


圖 30 池化過程示意圖

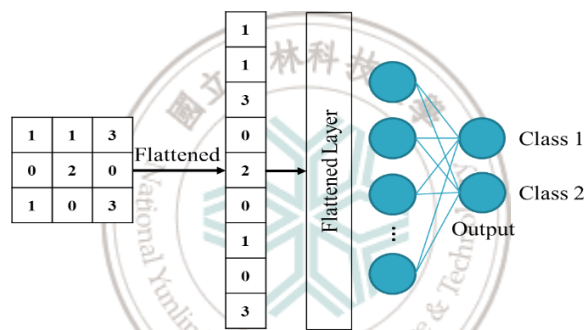


圖 31 全連接層示意圖

4.5.2 應用 CNN 之 Inception-ResNet 架構辨識內視鏡影像

4.5.2.1 CNN 介紹

卷積層

卷積層為從輸入之影像進行特徵擷取的特徵擷取層，其目的在於擷取輸入影像之特徵。卷積通過使用輸入之數據的陣列學習影像特徵來保留像素之間的關係，為一種數學矩陣中的運算，在運算中應用 $N \times N$ 尺寸之卷積核透過滑動窗格的方式進行計算，在計算後會取得該特徵矩陣之特徵圖 (Feature Map)，接著將特徵圖透過激活函數 (Activation Function) 進行非線性轉換，轉為非線性的目的在於讓模型學習更合適於輸入之影像特徵，通過卷積層的過程如圖 29 所示。

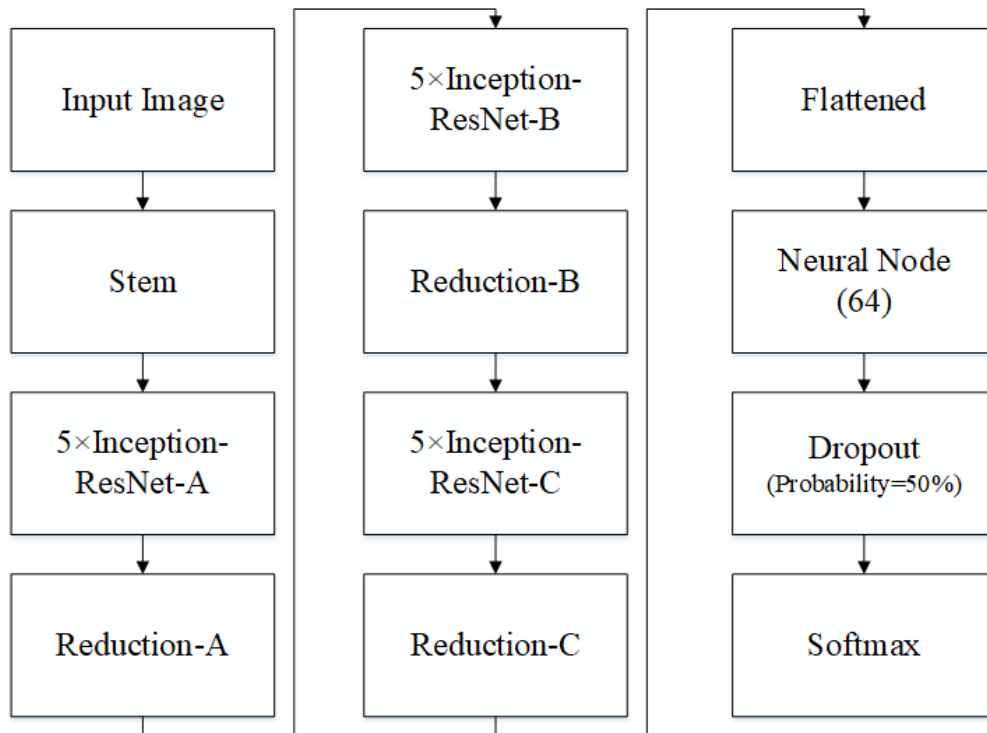


圖 32 Inception-ResNet 網路架構示意圖[56]

池化層(Pooling Layer)

其目的在於從卷積層中取得之特徵圖進行降維的動作，將特徵圖的尺寸縮小，保留特徵圖中重要的特徵，其運算的方式與卷積層之計算過程相似，池化層採用 $N \times N$ 之卷積核經由滑動窗格的方式進行運算，池化層採取的常見方式主要分為最大值池化(Max Pooling)、最小值池化(Min Pooling)及平均值池化(Average Pooling)，如圖 30 所示，藉由不同之卷積層的為最後輸出之特徵值不會經由激活函數進行非線性轉換運算。

全連接層(Fully Connected Layer)

全連接層是學習卷積層輸出中所取得之非線性重要特徵組合最常用的方法，透過全連接層可以學習該影像中所有可能存在的非線性函數，通常將卷積層輸出之特徵進行攤平(Flatten)的動作，作為適合多層感知器能夠輸入的形式，卷積層輸出之特徵被輸入至前饋類神經網路進行特徵之學習，並透過反向傳播的方式應用於訓練的每一次迭代，於每一次迭代中進行權重的更新，在進行訓練的過程中，該模型能夠區分對該影像中的主要特徵及對影像非必要特徵，最後通過常用之激活函數 Softmax 或 Sigmoid 計算出各類別輸出之機率，機率越高則象徵該影像所代表之標籤。全連接層如圖 31 所示。

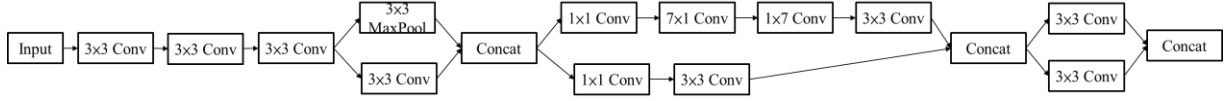


圖 33 Stem 示意圖[56]

4.5.2.2 Inception-ResNet 主要架構

本研究使用 Inception-ResNet 作為訓練之基礎模型，如圖 32 為 Inception-ResNet 之模型架構，以下將針對架構中之 Stem、Reduction 模組及 Inception-ResNet 模組進行介紹。

(1) Stem[56]

Stem 為用於擷取本論文中內視鏡影像中特徵訊息之結構，此結構透過池化層與卷積層的併行及不對稱之卷積核，在減少參數量的同時，也不會丟失網路中大量特徵訊息，此外也透過 1×1 卷積核來減少維度。Stem 架構如圖 33 所示。

(2) Inception-ResNet 模組[56]

Inception-ResNet 模組由 Inception[57]及 ResNet[27]兩種模型之概念所組成，Inception 的概念為將 3 種卷積核跟一個 3×3 之最大池化疊在一起，其堆疊在一起的原因為可以擷取影像中更多細節之特徵，而 1×1 卷積核的壓縮可以減少影像的維度及減少訓練之參數，這樣做的意義主要是為了控制輸出通道的深度，並同時能增加模型的非線性；ResNet 的概念為藉由恆等對映(identity mapping)的概念對網路架構進行改善，如公式(6)所示。

$$H(x) = F(x) + x \quad (6)$$

公式中 $H(x)$ 則為經過網路學習後之輸出值， $F(x)$ 為網路需學習之權重， x 表示為原始輸入之 x 。在訓練的過程當中，網路會判斷該輸入之網路層是否為多餘之網路層，若判斷為冗餘層，即將輸入與輸出值恆等皆為 x ，亦即經過捷徑連接(shortcut connections)此捷徑，若判斷該網路層不為冗餘層，則學習 $F(x)$ 並輸出 $H(x)$ ，其每次保留輸入層之值可以使網路在反向傳播時保證至少會有一個 1 存在，殘差網路進行反向傳播公式如 (7)，進而降低梯度問題的可能性，如圖 34 為殘差學習示意圖。

$$\frac{\partial \varepsilon}{\partial x_i} = \frac{\partial \varepsilon}{\partial x_L} \frac{\partial x_L}{\partial x_i} = \frac{\partial \varepsilon}{\partial x_L} \left(1 + \frac{\partial}{\partial x_i} \sum_{i=1}^{L-1} F(x_i, W_i) \right) \quad (7)$$

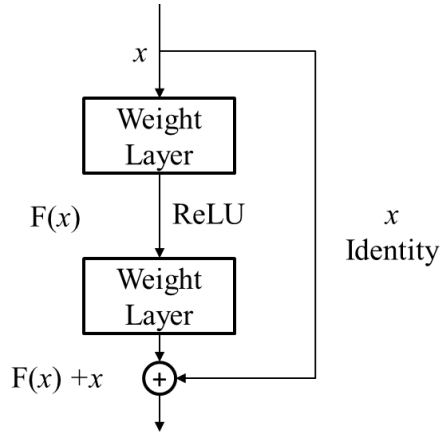


圖 34 ResNet 示意圖[27]

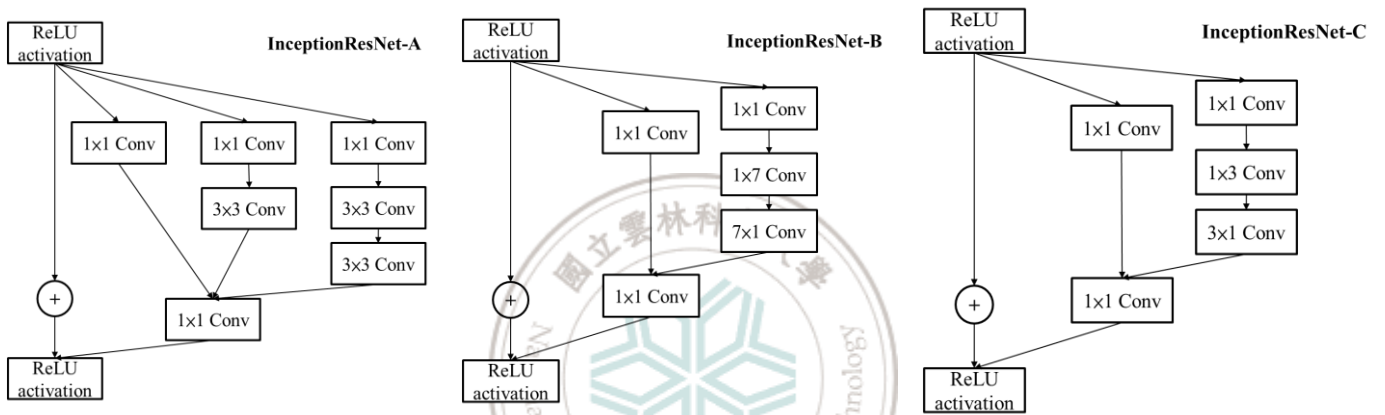


圖 35 Inception-Resnet 各模組示意圖[56]

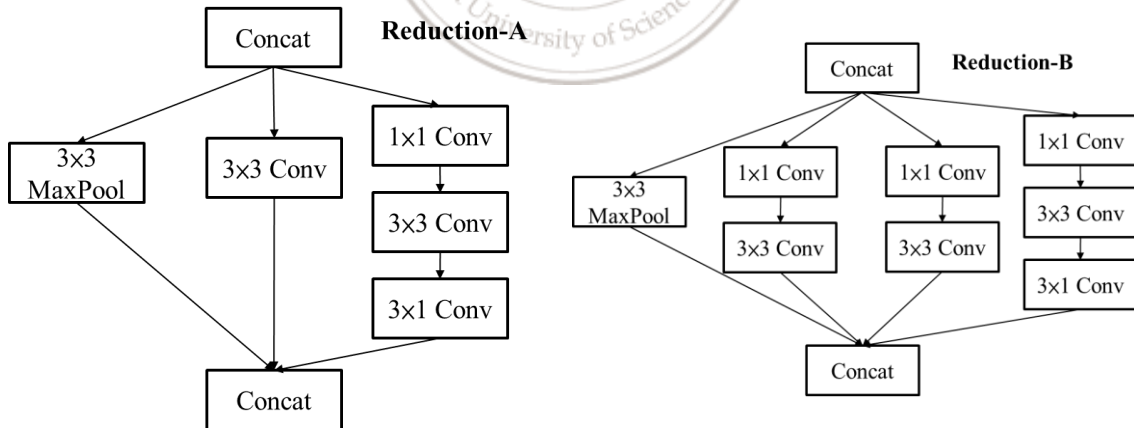


圖 36 Reduction 模組示意圖[56]

Inception-Resnet 將上述概念結合後，於本論文訓練內視鏡影像辨識模型中可以使模型在隨著網路增加時避免梯度消失的問題也能透過 Inception 的分支效果擷取內視鏡影像中更多特徵訊息，並將各模組串接，透過串接的方法也使得網路層數增加進而讓網路更能學習影像中之特徵。Inception-Resnet 各模組示意圖如圖 35 所示。

(3) Reduction 模組[56]

由於隨著 Inception-ResNet 模組的疊加，使得欲訓練之內視鏡辨識網路架構越來越深，因此於 Reduction 模組中將最後 3 個 3×3 之卷積核步長設置為 2 可以有效的降低特徵圖的尺寸，同時 Reduction 的架構當中也是使用併行、不對稱卷積層及 1×1 卷積核來降低參數量。Reduction 模組架構示意圖如圖 36 所示。

4.5.3 CNN 模型訓練

本研究使用模型 Inception-ResNet 之模型訓練過程如圖 37：

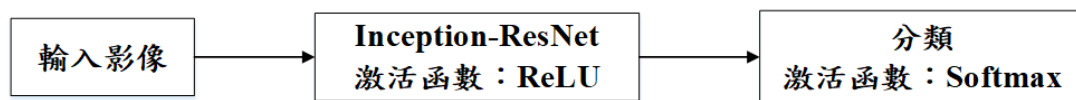


圖 37 模型訓練過程

輸入影像

本研究之輸入影像為原始色影像、不同色彩空間影像及不同亮度影像各別輸入進行訓練，示意圖如圖 38 所示。

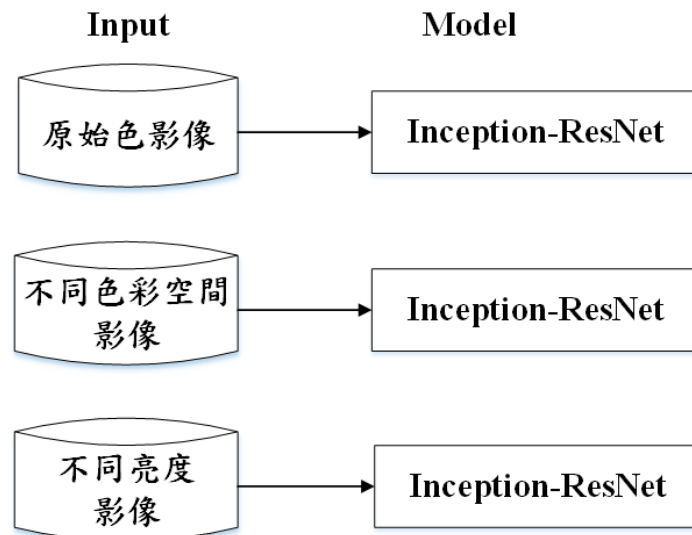


圖 38 各影像輸入模型示意圖

Inception-ResNet

於 Inception-ResNet 模型中進行內視鏡影像特徵擷取，於每個卷積核計算後之特徵圖使用激活函數 ReLU(Rectified Linear Unit)非線性轉換，而最後分類的部份使用激活函數 Softmax 輸出各類別之機率，並於最後使用損失函數使預測值更貼近於真實值。

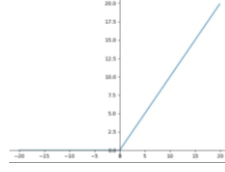


圖 39 ReLU 函數

本論文使用 ReLU 作為網路之激活函數，為了使卷積核中內視鏡影像之特徵值轉換為非線性的特徵，讓神經網路在訓練的過程當中學習非線性的輸出，同時也使訓練具有效性，進而降低過度擬合情況的發生，因此於每一層卷積層輸出後透過 ReLU 函數進行非線性轉換，圖 39 為 ReLU 之函數示意圖及式(8)為其公式，卷積運算過後之矩陣中每一個像素點之值若大於 0，則輸出該值大小，若小於 0，則將該像素點之值輸出為 0，經由運算後，輸出之特徵圖有部分之像素值輸出為 0，其減少了特徵值之間的相依性，並且讓網路具有稀疏性降低過度擬合的問題。

$$f(x) = \begin{cases} 0, & \text{if } x < 0 \\ x, & \text{otherwise} \end{cases} \quad (8)$$

以 Softmax 作為最後分類激活函數，於分類問題中最常使用的函數，本論文輸出為二值化類別，即將影像分成兩類(陰性、陽性)，所有類別之預測值經過此函數運算後會限制在區間 0 至 1 之機率值，且總和為 1。在最後一層網路層時，得到陰性及陽性兩類別之機率值， z_i 為陰性及陽性類別之值，輸出其類別機率值(e^{z_i})，並且所有網路輸出類別之機率值總合($\sum_{k=1}^K e^{z_k}$)為 1，其函數公式如式(9)。

$$f(z)_i = \frac{e^{z_i}}{\sum_{k=1}^K e^{z_k}}, \text{ for } i = 1, \dots, K \quad (9)$$

以 Categorical Cross-Entropy (CE) 作為損失函數，其函數主要功能為計算實際標記結果(Ground Truth)與預測分類結果(Predict Result)之間所產生出來之偏差值，根據此偏差值來去判斷模型學習後優劣之重要依據，公式如式(10)所示， n 為樣本數， m 為類別數， y_i 為真實值， \hat{y}_i 為預測值。

$$loss = -\sum_{i=1}^n y_i \cdot \log \hat{y}_i \quad (10)$$

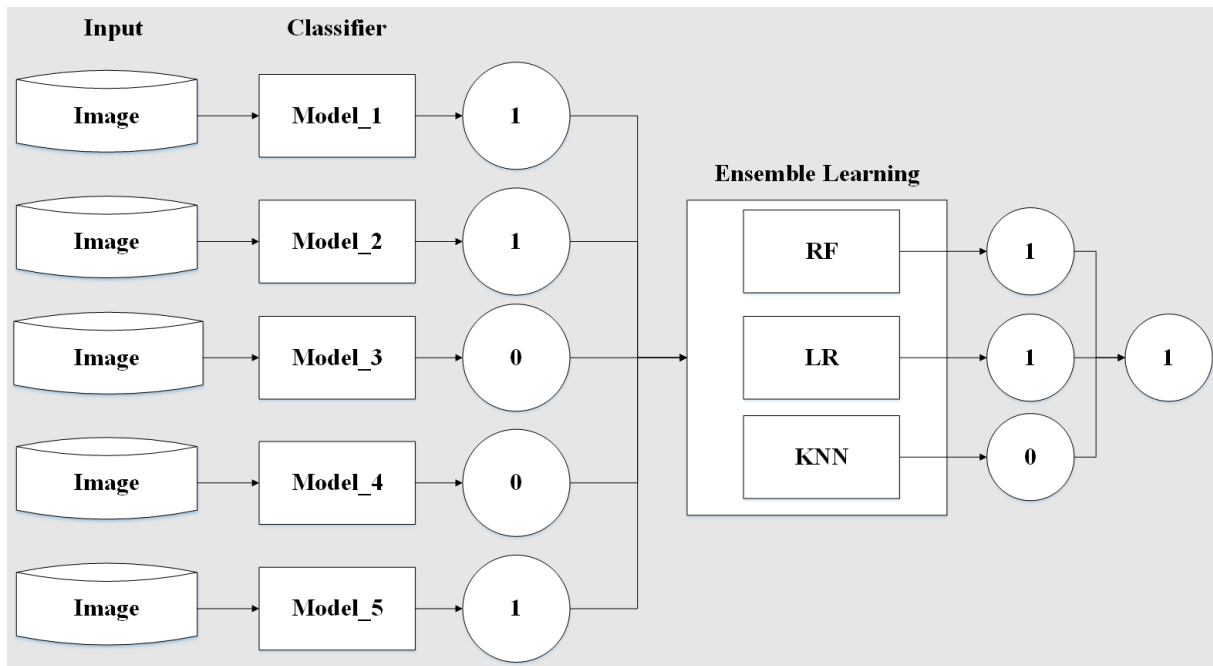


圖 40 集成學習之概念圖

4.5.4 集成學習框架

在本小節集成框架的部分將會以 Stacking 的方式對各模型進行集成，以圖 40 為例，若本論文設計 5 個 CNN 模型，將影像各別輸入至 5 個分類器模型當中，各分類器模型會輸出影像之類別結果，亦即 3 個 1 及 2 個 0 之輸出結果，接著集成學習會將分類器模型之輸出結果再輸入之集成學習模型內，集成學習內之模型會重新根據這些分類器模型之輸出結果再重新訓練，而本研究所使用的集成學習之模型將會以機器學習之分類器來呈現，最後會依據此 3 種分類器之結果再藉由投票法則得到最後的結果。

4.5.5 績效衡量

本研究使用混淆矩陣如表 5 做為績效衡量，True Positive (TP)：真實鼻咽腫塊為陽性且預測鼻咽腫塊為陽性的張數，False Negative (FN)：真實鼻咽腫塊為陽性卻預測鼻咽腫塊為陰性的張數，False Positive (FP)：真實鼻咽腫塊為陰性卻預測有鼻咽腫塊為陽性的張數，True Negative (TN)：真實鼻咽腫塊為陰性且預測鼻咽腫塊為陰性的張數，利用 TP、TN、FN 及 FP 來評估模型預測的正確率(Accuracy)、精確率(Precision)、召回率(Recall)、F-score。

表 5 混淆矩陣

		Real	
		Positive	Negative
Prediction	Positive	Ture Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	Ture Negative (TN)

正確率(Accuracy)

指模型整體判斷的正確率，其越高表示模型能越精確的分類出該圖像為陰性或陽性，正確率公式如式(11)所示。

$$Accuracy = \frac{TP + TN}{(TP + TN + FP + FN)} \quad (11)$$

召回率(Recall)

影像資料中的「所有目標」但「正確分類為目標」的比例。(所有鼻咽內視鏡影像中，被模型辨識出有鼻咽腫塊的比例)，召回率公式如式(12)所示。

$$Recall = \frac{TP}{TP + FN} \quad (12)$$

精確率(Precision)

「所有被檢測為目標」但「正確分類為目標」的比例。(所有模型預測鼻咽癌照片中，真的有鼻咽腫塊的比例)，精確率公式如式(13)所示。

$$Precision = \frac{TP}{TP + FP} \quad (13)$$

F-score(F1)

精確率及召回率兩者調和的平均數，為一個比較概略的指標來看這個模型的表現。

由公式(7)可知分母為真實鼻咽腫塊為陰性的總張數，分子為真實鼻咽腫塊為陰性且預測鼻咽腫塊為陰性的張數，因此計算出來的數值越高代表在預測鼻咽腫塊為陰性的影像準確率越高，數值越低則反之，F-score 公式如式(14)所示。

$$F - score = \frac{2 \times precision \times recall}{(precision + recall)} \quad (14)$$



第五章 基於 YOLO 模型之做法

第五章為介紹 YOLO 方法，在 Offline 運作部分，第一階段為內視鏡影像前處理的階段，擷取 ROI 影像區塊，並去除內視鏡影像中強光之雜訊取得原始色影像，最後進行資料增強，而 YOLO 輸入的部分將原始色影像資料集劃分為 N 個訓練資料集並進行模型訓練。第二階段為 YOLO 模型嘗試擷取各影像資料集之特徵，透過模型訓練後輸出物件之類別及物件位置。第三階段中將採用 4.5.4 小節中集成學習框架將 YOLO 之 N 個輸出之結果輸入至該框架進行集成，優化整體模型框架。Online 運作部分，則會依據 Offline 中 YOLO 所挑選之特徵值應用於實際鼻咽內視鏡影像當中。Offline、Online 流程如圖 41 所示。而詳細的 YOLO 集成框架部分於 5.2.3 進行統整。

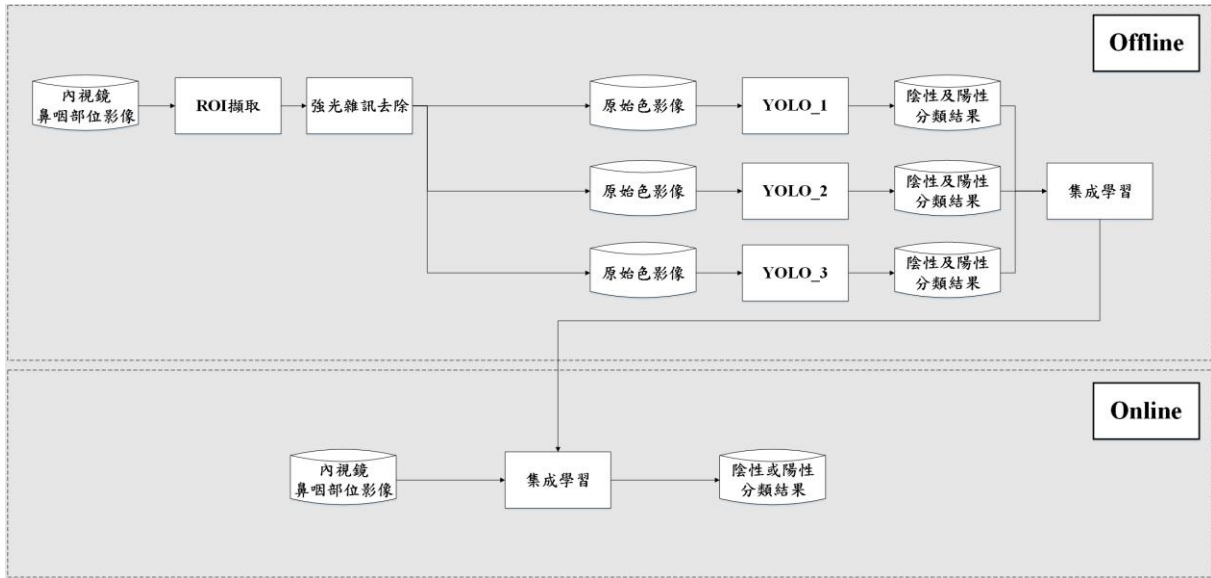


圖 41 YOLO 研究流程圖

5.1 YOLO 內視鏡影像前處理

YOLO 與 CNN 模型不同之處在於 YOLO 需要輸入欲預測之物件位置及物件類別，因此需特別使用 LabelImg 開放原始碼標籤程式為 4.1、4.2 中擷取 ROI 及強光雜訊去除後之原始色影像資料進行標籤，標籤物件後會呈現縱向之位置值(X)、標籤物件橫向之位置值(Y)、標籤物件之寬(Width, W)、標籤物件之高(Height, H)、影像類別(0：陰性，1：陽性)，如表 6 所示。陰性及陽性標籤如圖 42、圖 43 所示。



圖 42 鼻咽標籤陰性範例圖

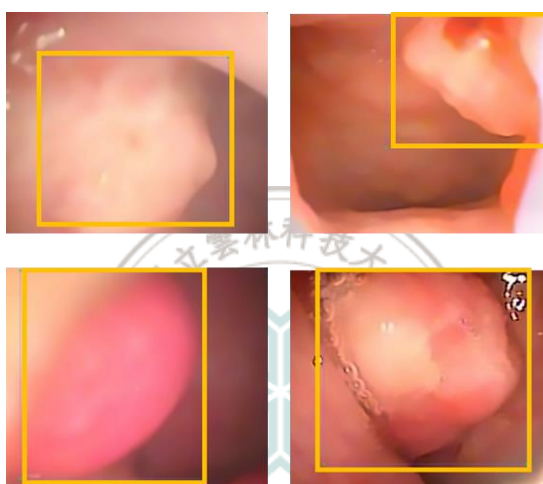


圖 43 鼻咽標籤陽性範例圖

表 6 YOLO 內視鏡鼻咽部位標籤

影像名稱	X	Y	W	H	Class
0475311_0810211420_I00064468876	0.5	0.5	0.994	0.993	0
0547606_0704090615_I00052501407	0.5	0.5	0.994	0.993	0
0208952_0603300639_I00045164832	0.5	0.5	0.994	0.993	0
...
0045289_0510172120_I00041884860	0.672	0.316	0.616	0.626	1
0009911_0804032249_I00060161229	0.516	0.495	0.961	0.983	1
7961601_0512071474_I00042902874	0.494	0.588	0.711	0.756	1

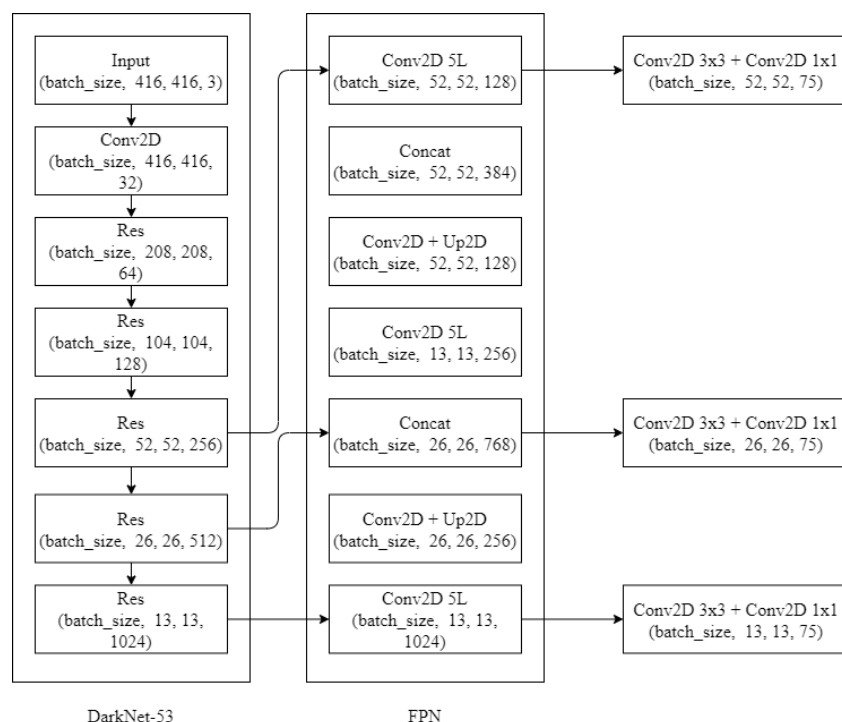


圖 44 YOLOv3 網路架構圖[51]

5.2 建立基礎影像辨識之 YOLO 模型

由於在 4.2.1 中介紹 CNN 中卷積層、池化層、全連接層之運作方式，本章節僅說明 YOLO 運作方式。

5.2.1 應用 YOLO 辨識內視鏡影像

主要架構

本研究使用 YOLOv3 作為第二個訓練之模型，圖 44 為 YOLOv3 之模型網路架構圖，YOLOv3 為由 DarkNet-53 與 FPN 所結合而成的，以下將針對架構中 DarkNet-53、FPN 進行介紹。

(1) DarkNet-53[51]

DarkNet-53 為 YOLOv3 卷積神經網路中之骨架，其主要由卷積核 1×1 尺寸之卷積及殘差網路中捷徑連接之概念結構所組成，總共使用 53 層卷積層建構，卷積層加深以後能更有效的對內視鏡影像擷取特徵。

(2) FPN[51]

由於在卷積運算的過程中隨著層級越高，遺失的特徵也會隨著增高，為了解決此問

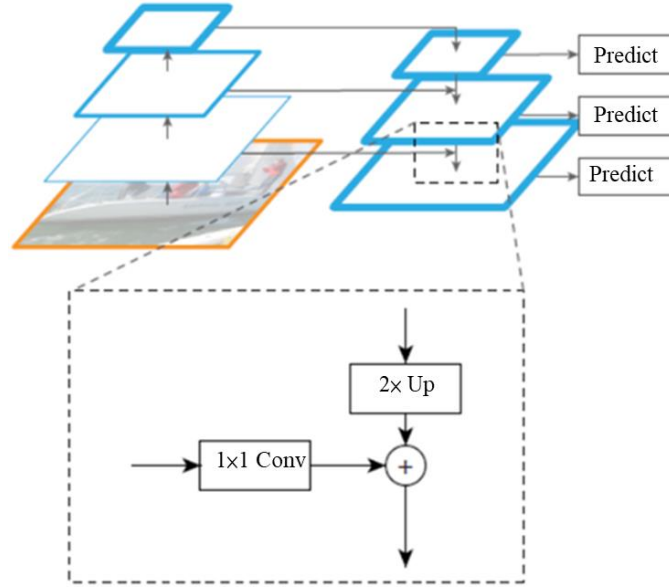


圖 45 FPN 架構示意圖[39]

題，FPN 藉由從低層到高層中的路徑(Top-Down Pathway)與橫向連結(Lateral Connection)方法，意即將從相同層級卷積層之特徵傳送到預測之卷積層，透過此方法可以使低層的內視鏡影像中特徵訊息與高層內視鏡影像中特徵特徵訊息能更有效的融合進而增加內視鏡影像之特徵訊息，造就 FPN 對於擷取影像中細小特徵更有優勢，圖 45 為 FPN 架構示意圖。

在 YOLOv3 中之 FPN 使用三種不同尺寸對內視鏡影像進行特徵擷取，分別為 13×13 、 26×26 和 52×52 ，其中 26×26 及 52×52 是採用上採樣的方式從原本 13×13 尺寸逐漸上升的，而各別尺寸 13×13 負責內視鏡影像中較大之物件、 26×26 負責內視鏡影像中中等之物件、 52×52 負責內視鏡影像中較小之物件，每個尺寸會預測出三個邊界框，其邊界框預測公式如式(15)到(18)所示，公式中 t_x, t_y, t_w, t_h 為網路中偏移量和尺寸縮放， b_x, b_y, b_w, b_h 為最後邊界框輸出結果。

$$b_x = \sigma(t_x) + c_x \quad (15)$$

$$b_y = \sigma(t_y) + c_y \quad (16)$$

$$b_w = p_w e^{t_w} \quad (17)$$

$$b_h = p_h e^{t_h} \quad (18)$$

5.2.2 YOLO 模型訓練

本研究使用 YOLOv3 基礎模型之模型訓練過程如圖 46 所示：

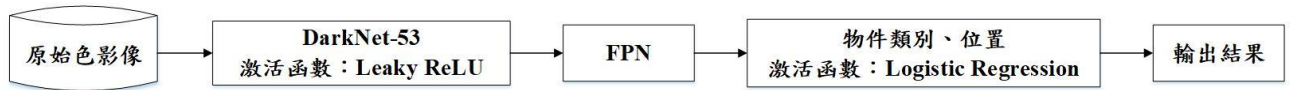


圖 46 YOLOv3 訓練過程

輸入影像及標籤之物件寬、高、位置

輸入原始色影像 299×299 尺寸之內視鏡影像與已標籤之內視鏡影像中物件 X、Y、寬、高值，示意圖如圖 47 所示。

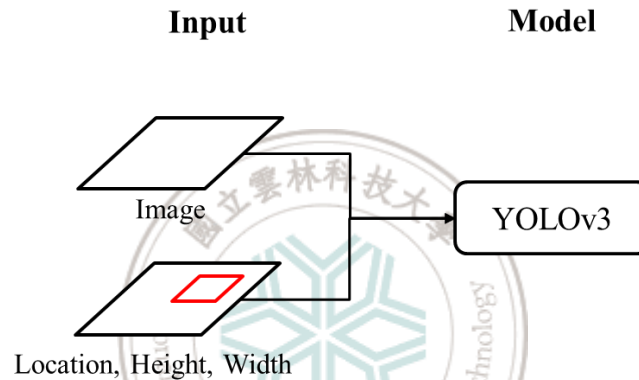


圖 47 影像輸入模型示意圖

DarkNet-53

影像於 DarkNet-53 模型中透過 53 層卷積層分別對內視鏡影像進行特徵擷取，於各別卷積層擷取內視鏡影像之特徵圖分別使用激活函數 Leaky ReLU 進行非線性轉換。

(1) 激活函數

A. Leaky ReLU

由於 ReLU 會因為梯度過大導致參數更新過之後，神經元不再對任何數據有激活現象，為了解決此問題，Leaky ReLU 給所有小於 0 之值賦予一個非零之斜率，圖 48 為 Leaky ReLU 之函數示意圖及式(19)為其公式。

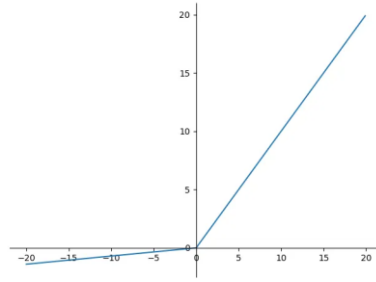


圖 48 Leaky ReLU 函數

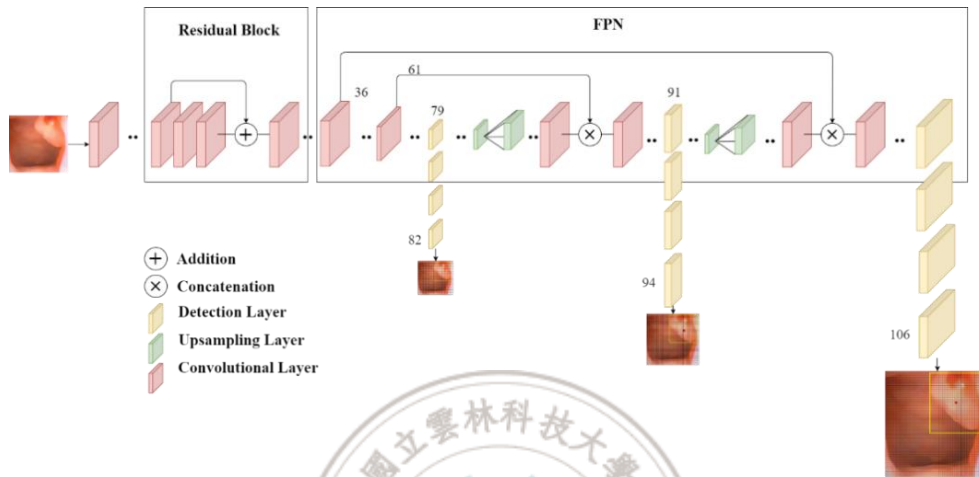


圖 49 FPN 輸出結果示意圖

$$f(x) = \begin{cases} 0.01x, & \text{if } x < 0 \\ x, & \text{otherwise} \end{cases}$$

(19)

FPN

FPN 會從 53 層卷積層中擷取之內視鏡影像特徵中藉由從低層到高層中的路徑與橫向連結方法，從相同層級卷積層之內視鏡影像特徵傳送到欲預測之卷積層，最後輸出 13×13、26×26 和 52×52 尺寸之結果，圖 49 為三種尺寸輸出之結果。

類別、物件位置

最後分類的部份將 FPN 三種尺寸之內視鏡影像輸出影像結果使用 Logistic Regression(LR)對每個框中影像進行分類預測，接著透過 NMS 輸出最好的內視鏡影像邊界框及其類別，最後透過 Binary Cross-Entropy(BCE)對 YOLOv3 中網路權重的更新，使最後預測框及分類值能更貼近於真實值，LR、NMS、BCE 公式如下所示。

在分類的階段，我們使用分類函數為 LR，其為了解決過去 Softmax 只能對應至一個類別的問題，採用了 LR 作為分類函數，在 YOLOv3 中，LR 只會對 Ground Truth 中 IoU 最大之邊界框計算分類、迴歸之值，LR 公式如式(20)所示，其算出參數後藉由 sigmoid

進行非線性轉換，輸出最後的值，sigmoid 公式如式(21)所示。

$$y = \text{sigmoid}(w^T x + b) = \sigma(w^T x + b) \quad (20)$$

$$\text{sigmoid}(z) = \sigma(z) = \frac{1}{(1 + e^{-z})} \quad (21)$$

在損失函數中，我們使用的損失函數為 Binary Cross-Entropy (BCE)，其函數主要功能如 CE 一樣為計算實際標記結果(Ground Truth)與預測分類結果(Predict Result)之間所產生出來之偏差值，依據此偏差值判斷模型學習後之依據，於 YOLOv3 中透過 BCE 計算物件置信度及分類預測之損失來更新網路中權重使物件邊界框及類別更貼近於實際標記結果，公式如式(22)所示， n 為樣本數， y_i 為真實值， y_i 為預測值。

$$\text{loss} = -\sum_{i=1}^n y_i \log y_i + (1 - y_i) \log(1 - y_i) \quad (22)$$

5.2.3 集成學習框架

取得經影像前處理後之原始色影像資料級，將原始色影像資料集隨機抽取 N 個樣本作為 N 個訓練集，各 N 個訓練集會各別輸入至各別之 YOLOv3 模型進行訓練，模型訓練後，輸入之影像會由各已訓練之 N 個 YOLOv3 模型輸出各別分類結果，最後透過集成學習中之機器學習分類器進行預測，各分類器預測之結果會基於投票法獲取最後分類結果。以圖 50 為例，若本論文設計從原始色資料集隨機抽取 3 個樣本作為訓練集對各別 YOLOv3 模型進行訓練，未來將新的原始色影像再輸入至已訓練模型中預測分類值，若 3 個模型中預測出兩個為 1 之類別及一個為 0 之類別，則將此 3 個分類結果再輸入集成學習模型當中再進行預測，最後依據集成學習模型中三個分類器之結果再透過投票法則輸出 1 之結果。

5.2.4 績效衡量

YOLOv3 也同樣使用如表 5 混淆矩陣做為績效衡量，利用 TP、TN、FN 及 FP 來評估模型預測的正確率(Accuracy)、精確率(Precision)、召回率(Recall)、F-score，正確率等公式已於 4.2.3 節中說明，所以我們不在此節列出，以下為 YOLO 專屬。

IoU(Intersection over Union)

由 Paul Jaccard 於 1901 年提出，亦可稱為 Jaccard 相似係數，它是為了比較真實標記(Ground Truth)與預測分割結果(Predict Result)之邊界框相似性的統計量，其公式如式(23)所示。

$$IoU = \frac{ground\ truth \cap predict\ result}{ground\ truth \cup predict\ result} \quad (23)$$



第六章 實驗模擬

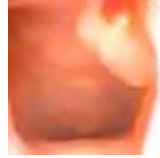
本論文使用由 S 醫院提供之鼻咽部位影像資料集，並使用本論文提出之方法來驗證其有效性，以下將分為六個部分，(1)資料集與實驗參數介紹，(2)CNN、YOLO 模型之有效性，(3)模型可視化分析探討，(4)集成學習之模型選擇，(5)模型之績效與評估(6)模型質化合理性說明。本實驗皆由 Python 完成，並操作在 Intel® Core™ i7-9700KF 的 CPU、Nvidia Geforce RTX 2080 Ti 11GB 的 GPU、搭載 DDR4 16GB 記憶體以及 Windows 10 的作業軟體來完成。

6.1 資料集與實驗參數介紹

本論文使用之影像資料集為由 S 醫院提供，其影像資料集由該 S 醫院醫師於診斷病患時所擷取之影像。首先將醫生以手稿的方式圈選鼻咽腫塊患部位置，接著以電子化方式整理出欲訓練之資料集，最終影像總筆數為 234 筆影像，其中陽性總筆數為 127 筆影像，陰性總筆數為 107 筆影像，並將資料拆分為比例為總筆數 0.85 的訓練集與總筆數 0.15 的測試集，最終訓練集樣本數為 198 筆、測試集樣本數為 36 筆。本研究將影像輸入至本實驗中之模型來辨識該影像是否有腫塊。本論文將所有影像大小長、寬設置為 299×299 ，並將影像整理為原始色影像、不同色彩空間影像、不同亮度影像及 YOLO 標記之原始色影像，其中不同亮度影像之參數設置以 0.8 至 2.0 的 10 個平均區間參數，分別為 0.8、0.93、1.06、1.2、1.33、1.47、1.6、1.73、1.87、2.0，亮度參數愈高則亮度愈強，亮度參數愈低則反之，圖 51 為不同參數亮度示意圖，表 7 為影像轉換、亮度參數及訓練模型之列表。接著由於總樣本數僅有 234 筆還不足以使模型於訓練時學習陽性影像中腫塊之區塊，同時也為了避免過擬和的情況發生，因此本論文透過資料增強之方法來增加樣本數使模型學習更多樣本，資料增強在模型學習當中會將影像藉由垂直翻轉、水平翻轉、隨機旋轉、隨機縮放隨機進而達到產生不同樣本之效果。在 CNN 輸入模型的部分，將原始色影像、不同色彩影像、不同亮度之訓練集各自輸入至 CNN 網路模型當中，透過 CNN 能擷取較好的特徵值特性，最後透過攤平及神經元學習 CNN 中之特徵，預測出該影像之結果；在 YOLO 輸入模型的部分，則是將僅將原始色影像訓練集之影像及標記



(a)參數 0.8



(b)參數 1.2



(c)參數 1.8

圖 50 不同亮度示意圖(a)參數 0.8、(b) 參數 1.2、(c) 參數 1.8

輸入至 YOLO 模型當中，透過 YOLO 中 53 層卷積層擷取不同影像中特徵，並經由 FPN 獲取不同尺寸之影像結果，最後通過 LR 對每個框中影像進行影像之分類預測，同時為了確保本實驗之嚴謹性，將每一種模型進行交叉驗證，以確保模型之可行性，並經由集成學習來優化整體方法之成效，最後為了確保本論文之有效性，使用了分類常見之模型效能評估方式，包含了混淆矩陣、正確率、召回率、精確率及 F-score 作為模型績效驗證指標。

表 7 影像轉換之參數及訓練之模型

影像類別	影像轉換之參數	訓練之模型
原始色影像	無	CNN、YOLO
不同色彩空間	熱區圖、HSV	CNN
不同亮度	0.8、0.93、1.06、1.2、1.33、1.47、1.6、1.73、1.87、2.0	CNN

6.2 CNN、YOLO 模型之有效性

本小節進行模型有效性之驗證，分別去查看 CNN 之不同色彩空間、不同亮度模型及 YOLO 之績效，下表 8、9、10、11 為各別 CNN 及 YOLO 模型辨識的混淆矩陣及績效結果，在訓練集的部分，絕大多數的模型準確度及 F-score 都可以超過 0.9，僅 HSV 及熱區圖未超過 0.9，在測試集的部分可以看到，大部分模型的準確度及 F-score 也幾乎可以大於 0.85，僅有少部分的模型接近於 0.8，也可以發現的是亮度在參數越高的情況下模型的績效上較沒有那麼好，最後在有效性的部分，為了確保其嚴謹性將透過 5 次的交叉驗證來驗證各模型之可行性。

表 8 各模型訓練集混淆矩陣

原始色			
Training population(N=198)		Real	
Predict	Positive	Positive	Negative
	Negative	90	13
熱區圖			
Training population(N=198)		Real	
Predict	Positive	Positive	Negative
	Negative	83	20
HSV			
Training population(N=198)		Real	
Predict	Positive	Positive	Negative
	Negative	92	11
亮度_0.8			
Training population(N=198)		Real	
Predict	Positive	Positive	Negative
	Negative	93	10
亮度_0.9333			
Training population(N=198)		Real	
Predict	Positive	Positive	Negative
	Negative	92	11
亮度_1.0667			
Training population(N=198)		Real	
Predict	Positive	Positive	Negative
	Negative	97	6

表 8 各模型訓練集混淆矩陣(續)

亮度_1.2			
Training population(N=198)		Real	
Predict	Positive	Positive	Negative
	Negative	94	9
亮度_1.3333			
Training population(N=198)		Real	
Predict	Positive	Positive	Negative
	Negative	92	11
亮度_1.4667			
Training population(N=198)		Real	
Predict	Positive	Positive	Negative
	Negative	94	9
亮度_1.6			
Training population(N=198)		Real	
Predict	Positive	Positive	Negative
	Negative	95	8
亮度_1.7333			
Training population(N=198)		Real	
Predict	Positive	Positive	Negative
	Negative	92	11
亮度_1.8667			
Training population(N=198)		Real	
Predict	Positive	Positive	Negative
	Negative	92	11
亮度_2.0			
Training population(N=198)		Real	
Predict	Positive	Positive	Negative
	Negative	89	14

表 8 各模型訓練集混淆矩陣(續)

YOLO_1			
Training population(N=198)		Real	
		Positive	Negative
Predict	Positive	91	12
	Negative	5	90
YOLO_2			
Training population(N=198)		Real	
		Positive	Negative
Predict	Positive	86	17
	Negative	9	86
YOLO_3			
Training population(N=198)		Real	
		Positive	Negative
Predict	Positive	102	1
	Negative	26	69
表 9 各模型測試集混淆矩陣			
原始色			
Testing population(N=36)		Real	
		Positive	Negative
Predict	Positive	23	1
	Negative	2	10
熱區圖			
Testing population(N=36)		Real	
		Positive	Negative
Predict	Positive	19	5
	Negative	3	9
HSV			
Testing population(N=36)		Real	
		Positive	Negative
Predict	Positive	18	6
	Negative	1	11
亮度_0.8			
Testing population(N=36)		Real	
		Positive	Negative
Predict	Positive	21	3
	Negative	1	11

表 9 各模型測試集混淆矩陣(續)

亮度_0.9333			
Testing population(N=36)		Real	
Predict	Positive	Positive	Negative
	Negative	20	4
亮度_1.0667			
Testing population(N=36)		Real	
Predict	Positive	Positive	Negative
	Negative	21	3
亮度_1.2			
Testing population(N=36)		Real	
Predict	Positive	Positive	Negative
	Negative	22	2
亮度_1.3333			
Testing population(N=36)		Real	
Predict	Positive	Positive	Negative
	Negative	1	11
亮度_1.4667			
Testing population(N=36)		Real	
Predict	Positive	Positive	Negative
	Negative	22	2
亮度_1.6			
Testing population(N=36)		Real	
Predict	Positive	Positive	Negative
	Negative	4	8
亮度_1.7333			
Testing population(N=36)		Real	
Predict	Positive	Positive	Negative
	Negative	21	3

表 9 各模型測試集混淆矩陣(續)

亮度_1.8667				
Testing population(N=36)		Real		
		Positive	Negative	
Predict	Positive	20	4	
	Negative	2	10	
亮度_2.0				
Testing population(N=36)		Real		
		Positive	Negative	
Predict	Positive	22	2	
	Negative	2	10	
YOLO_1				
Testing population(N=36)		Real		
		Positive	Negative	
Predict	Positive	21	3	
	Negative	2	10	
YOLO_2				
Testing population(N=36)		Real		
		Positive	Negative	
Predict	Positive	24	0	
	Negative	2	10	
YOLO_3				
Testing population(N=36)		Real		
		Positive	Negative	
Predict	Positive	22	2	
	Negative	3	9	

表 10 各模型訓練集實驗結果

Model	Accuracy	Precision	Recall	F1
Original	0.91	0.8738	0.9474	0.91
熱區圖	0.8384	0.8058	0.8737	0.8384
HSV	0.8687	0.8932	0.8598	0.8762
亮度_0.8	0.9141	0.9029	0.93	0.9163
亮度_0.9333	0.9242	0.8932	0.9583	0.9246
亮度_1.0667	0.9242	0.9126	0.94	0.9261
亮度_1.2	0.9495	0.9417	0.9604	0.951
亮度_1.3333	0.9293	0.9126	0.9495	0.9307
亮度_1.4667	0.9293	0.8932	0.9684	0.9293
亮度_1.6	0.9242	0.8932	0.9583	0.9246

表 10 各模型訓練集實驗結果(續)

亮度_1.7333	0.9242	0.9223	0.9314	0.9268
亮度_1.8667	0.9091	0.8641	0.957	0.9082
亮度_2.0	0.9293	0.8932	0.9684	0.9293
YOLO_1	0.9141	0.8835	0.9479	0.9146
YOLO_2	0.9444	0.835	0.9053	0.96
YOLO_3	0.8636	0.9903	0.7969	0.8831

表 11 各模型測試集實驗結果

Model	Accuracy	Precision	Recall	F1
Original	0.9167	0.9583	0.92	0.9388
熱區圖	0.8056	0.75	0.9474	0.8372
HSV	0.7778	0.7917	0.8636	0.826
亮度_0.8	0.8889	0.875	0.9545	0.913
亮度_0.9333	0.8611	0.8333	0.9524	0.8889
亮度_1.0667	0.9167	0.875	1.0	0.9333
亮度_1.2	0.9167	0.9167	0.9565	0.9362
亮度_1.3333	0.9167	0.9167	0.9565	0.9362
亮度_1.4667	0.8333	0.9167	0.8462	0.88
亮度_1.6	0.8611	0.9583	0.8519	0.902
亮度_1.7333	0.8333	0.875	0.875	0.875
亮度_1.8667	0.8333	0.8333	0.9091	0.8696
亮度_2.0	0.8889	0.9167	0.9167	0.9167
YOLO_1	0.8611	0.875	0.913	0.8936
YOLO_2	0.9444	1.0	0.923	0.96
YOLO_3	0.8611	0.9167	0.88	0.898

為了確保實驗之嚴謹性及模型之有效性，我們使用 K-fold Cross Validation 來驗證各個模型，並將影像資料集拆分成 5 等分進行 5 次的交叉驗證，以便於去觀察不同資料分布的情況下模型預測的狀況。表 12 為各模型在 5 次交叉驗證之準確度結果。

表 12 各模型交叉驗證實驗結果

模型	Acc.1	Acc.2	Acc.3	Acc.4	Acc.5	Mean
Original	0.8511	0.8511	0.8936	0.7872	0.7826	0.8331
熱區圖	0.7660	0.7447	0.7872	0.7872	0.8478	0.7866
HSV	0.7234	0.8085	0.8085	0.8298	0.8043	0.7949
亮度_0.8	0.7872	0.7660	0.8511	0.8085	0.8043	0.8034
亮度_0.9333	0.7447	0.7660	0.8936	0.7872	0.8043	0.7992
亮度_1.067	0.7872	0.7660	0.8723	0.8511	0.8043	0.8162
亮度_1.2	0.7234	0.7660	0.9149	0.8511	0.7826	0.8076
亮度_1.3333	0.8085	0.7660	0.9149	0.8298	0.8043	0.8247
亮度_1.4667	0.7234	0.8085	0.8936	0.7660	0.7826	0.7948

表 12 各模型交叉驗證實驗結果(續)

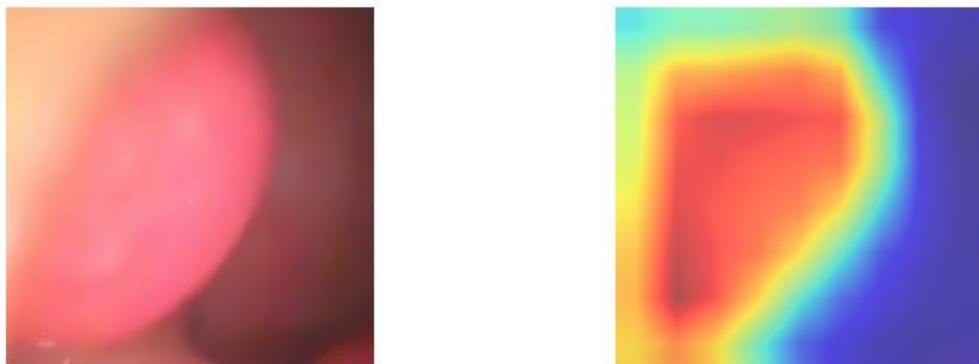
亮度_1.6	0.7447	0.8085	0.9149	0.7872	0.8043	0.8119
亮度_1.7333	0.6596	0.7872	0.9149	0.7021	0.7826	0.7693
亮度_1.8667	0.6596	0.7447	0.8723	0.6596	0.7609	0.7394
亮度_2.0	0.7021	0.7021	0.8511	0.7234	0.7391	0.7436
YOLO	0.8085	0.7447	0.8723	0.8511	0.8043	0.8162

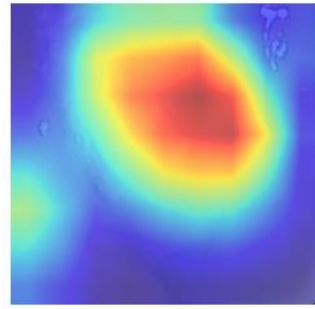
從 5 次交叉驗證實驗結果可以發現，在幾次的交叉驗證當中存在過擬和的情況發生，但是各個模型在 5 次的交叉驗證中之準確度都可以保持在接近 0.8 左右，甚至超過 0.8，因此可以驗證該影像資料集對於模型是可行的。

6.3 模型可視化分析探討

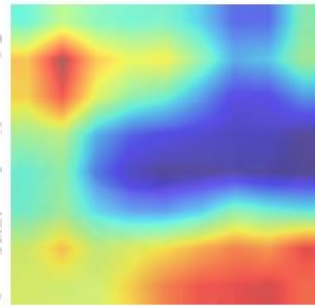
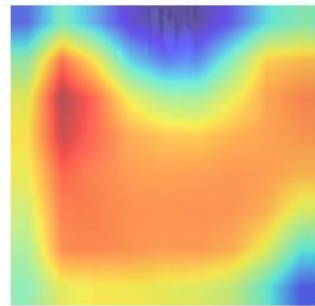
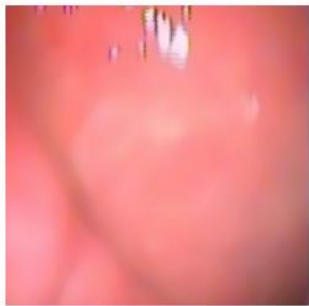
梯度加權類激活映射(Gradient-weighted Class Activation Mapping, Grad-CAM)[53]為常見用於 CNN 可視化之技術，以用來檢測該模型是否有專注於影像中欲關注之區塊，Grad-CAM 透過梯度之全局平均來計算權重，並藉由獲取每張特徵圖對應之權重進行加權總合，並透過熱區圖的方式來顯示該模型關注之區塊。在熱區圖的部分越紅的部分表示該模型越關注此區塊，若是以本實驗為例，期望 CNN 能關注於腫塊的區塊。

本小節使用 Grad-CAM 對 CNN 模型進行可視化，而 YOLO 的部分將會以邊界框及類別的方式進行可視化，以確保本實驗模型之有效性。由於模型過多的關係，因此不會將所有的模型進行可視化，僅以幾個可視化作為本小節之範例。



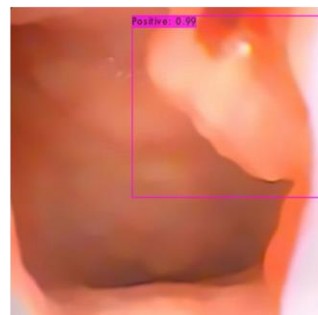


CNN 之 Grad-CAM 陽性(a)



CNN 之 Grad-CAM 陰性(d)

圖 51 (a)CNN 之 Grad-CAM 陽性視覺化結果(b)CNN 之 Grad-CAM 陰性視覺化結果



YOLO 陽性(a)

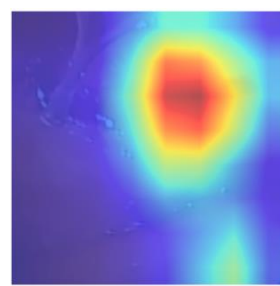


YOLO 陰性(b)

圖 52 (a)YOLO 陽性視覺化結果(b)YOLO 陰性視覺化結果

根據以上 CNN 之 Grad-CAM 視覺化及 YOLO 之視覺化之後的結果可以發現，CNN 之陽性影像在 Grad-CAM 熱區圖中都能有效的專注於患部區塊，在 YOLO 可視化當中也能有效的抓取陽性患部的區塊且也能有效的辨識陰性的部分。可視化圖例如圖 52、53 所示。

接著探討模型辨識錯誤的部分，首先為真實資料為陽性被辨識成陰性的部分，可以發現大部分陽性辨識成陰性的原因多為較不明顯的凸出部分或是其影像較為模糊的部分，或是腫塊區塊藏在較暗處以致於較難以預測該張影像為陽性，真實資料為陽性被辨識成陰性圖例如圖 54 所示；在真實資料為陰性被辨識成陽性的部分，其原因多為較明顯凸出部分類似於陽性腫塊輪廓的部分，或是該陰性影像平坦處相似於陽性影像平坦處導致影像在辨識上判定錯誤，真實資料為陰性被辨識成陽性圖例如圖 55 所示。

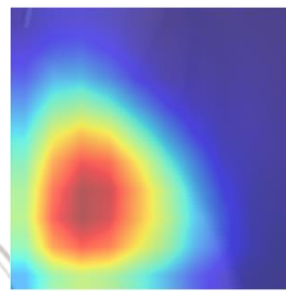


真實資料為陽性被辨識成陰性 CNN 之 Grad-CAM 視覺化結果 (a)

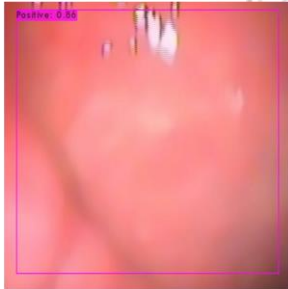


真實資料為陽性被辨識成陰性 YOLO 視覺化結果(b)

圖 53 (a) 真實資料為陽性被辨識成陰性 CNN 之 Grad-CAM 視覺化結果(b) 真實資料為陽性被辨識成陰性 YOLO 視覺化結果



真實資料為陰性被辨識成陽性 CNN 之 Grad-CAM 視覺化結果(a)



真實資料為陰性被辨識成陽性 YOLO 視覺化結果(b)

圖 54 (a) 真實資料為陰性被辨識成陽性 Grad-CAM 視覺化結果(b) 真實資料為陰性被辨識成陽性 YOLO 視覺化結果

最後總結在單一模型可能會導致模型辨識錯誤的幾個原因，本身影像較模糊、影像因為拍攝光源不穩定、影像患部區塊較不明顯及陰性影像所拍攝之患部輪廓與陽性相似。

6.4 集成學習之模型選擇

本小節在集成學習模型以 4 種不同的輸入端輸入至集成學習模型當中，第一個會以排列組合方式之 CNN、YOLO 輸出結果當作輸入，排列組合將從 CNN 有 13 種及 1 種

YOLO 模型共 16 種模型當中取 3 種到 5 種模型當作集成學習模型之輸入端，接著從取 3 到 5 種的模型當中觀察前 5 個最大的測試集準確度，並查看是否有重覆的特徵對於模型來說是較好的及取 3 到 5 種模型對於模型的績效是較好的，第二、三個以不同色彩空間、不同亮度之 CNN 輸出結果當作輸入，在 CNN 模型之輸入端部分會使用不同色彩空間影像、不同亮度影像作為 CNN 模型之輸入，第四個 YOLO 集成模型之輸入端部分則使用不同資料分布之原始色影像資料集作為 YOLO 之輸入，最後投票法皆由三種機器學習之分類器進行多數決的方式，而分類器會從常見的機器學習模型進行選擇，選擇的分類器分別有 LR、RF、KNN、Decision Tree(DT)、Support Vector Classifier(SVC)，在分類器最後的選擇上會以交叉驗證中最後平均準確度最高的三個當作集成學習之集成模型。

在不同色彩空間之 CNN 輸出結果作為集成學習中分類器之實驗結果顯示，LR、RF、KNN、DT、SVC 在經過 5 次的交叉驗證以後，LR、RF、KNN 的表現較佳，因此此三種分類器將會作為不同色彩空間集成學習之模型。表 13 為不同色彩空間集成學習分類器之交叉驗證實驗結果。

表 13 不同色彩空間集成學習分類器之交叉驗證實驗結果

模型	Acc.1	Acc.2	Acc.3	Acc.4	Acc.5	Mean
SVC	0.9362	0.8936	0.8511	0.9787	0.8913	0.9102
LR	0.9362	0.9787	0.8511	0.9787	0.8478	0.9185
RF	0.9362	0.9787	0.8511	0.9787	0.8478	0.9185
KNN	0.9362	0.9787	0.8511	0.9787	0.8478	0.9185
DT	0.9362	0.8936	0.8511	0.9787	0.8478	0.9015

在 10 種不同參數亮度 CNN 之輸出結果作為集成學習中分類器之實驗結果顯示，在實驗結果當中可以看到同樣 5 個模型在經過 5 次的交叉驗證以後，LR、RF、KNN 仍保有較佳的表現，因此以此三種分類器作為不同亮度集成學習之模型。表 14 為不同亮度集成學習分類器之交叉驗證實驗結果。

表 14 不同亮度集成學習分類器之交叉驗證實驗結果

模型	Acc.1	Acc.2	Acc.3	Acc.4	Acc.5	Mean
SVC	0.9362	0.9787	0.9149	0.9574	0.8913	0.9357
LR	0.9787	0.9574	0.9149	0.9787	0.9348	0.9529
RF	0.9574	0.9362	0.9149	0.9574	0.9348	0.9401
KNN	0.9787	0.9574	0.9362	0.9574	0.9565	0.9573
DT	0.9574	0.9362	0.9149	0.9149	0.9348	0.9316

在由 3 種不同切割之資料集影像分布之 YOLO 之輸出結果作為集成學習中分類器之實驗結果顯示，從實驗結果所呈現相同 5 個模型在經由 5 次的交叉驗證以後，LR、RF、KNN 在保持前三之平均準確度，因此仍以此三種分類器作為集成學習模型。表 15 為 YOLO 集成學習分類器之交叉驗證實驗結果。

表 15 YOLO 集成學習分類器之交叉驗證實驗結果

模型	Acc.1	Acc.2	Acc.3	Acc.4	Acc.5	Mean
SVC	0.9149	0.9149	0.8085	0.9362	0.7826	0.8714
LR	0.9362	0.9787	0.8511	0.9787	0.8478	0.9185
RF	0.9362	0.9787	0.8511	0.9787	0.8478	0.9185
KNN	0.9362	0.9787	0.8511	0.9787	0.8478	0.9185
DT	0.9362	0.8936	0.8511	0.9787	0.8478	0.9015

由於排列組合的組合數目過多，且從不同亮度、不同色彩空間及 3 種由 YOLO 集成所選擇之分類器可以看到，LR、RF、KNN 皆可以保持較好的績效，因此排列組合集成學習之分類器會繼續沿用 LR、RF、KNN 作為其集成學習模型的部分。

最後要選擇排列組合集成學習的輸入端特徵，首先排列組合的輸入將分別以 C_3^{16} 、 C_4^{16} 、 C_5^{16} 分別共有 560 種、1820 種、4368 種組合輸入至集成學習當中訓練，接著查看測試組準確度最大的前 5 組，表 16、17、18 為各排列組合輸入特徵及各別其實驗結果，從表中可以看到所有測試集的準確度皆為 0.9722，接續觀察訓練集的準確度，可以發現選擇 5 種模型可以使整體績效變得更好，因此選擇 5 種模型對於排列組合集成學習的輸入端為最佳的選擇，於表 16、17、18 集成學習輸入特徵整理之表 19 計數表可以發現，YOLO_3、原始色、YOLO_1、亮度_1.0667、亮度_1.7333 為各排列組合前 5 種最常出現之特徵，因此在輸入端特徵的部分將以此 5 種特徵輸入至集成學習當中。

表 16 排列組合 C_3^{16} 集成學習實驗結果

集成學習輸入特徵	Training Acc.	Testing Acc.
原始色、亮度_0.9333、YOLO_3	0.9596	0.9722
亮度_0.9333、YOLO_1、YOLO_3	0.9545	0.9722
熱區圖、YOLO_1、YOLO_3	0.9141	0.9722
熱區圖、YOLO_2、YOLO_3	0.9091	0.9722
熱區圖、HSV、YOLO_3	0.9040	0.9722

表 17 排列組合 C_4^{16} 集成學習實驗結果

集成學習輸入特徵	Training Acc.	Testing Acc.
原始色、亮度_1.0667、亮度_1.7333、YOLO_3	0.9697	0.9722
原始色、亮度_0.9333、亮度_1.7333、YOLO_3	0.9596	0.9722
原始色、亮度_0.9333、YOLO_1、YOLO_3	0.9596	0.9722
原始色、亮度_1.0667、亮度_1.8667、YOLO_3	0.9545	0.9722
HSV、亮度_0.8、YOLO_1、YOLO_3	0.9343	0.9722

表 18 排列組合 C_5^{16} 集成學習實驗結果

集成學習輸入特徵	Training Acc.	Testing Acc.
原始色、亮度_1.0667、亮度_1.6、YOLO_1、YOLO_3	0.9848	0.9722
原始色、亮度_1.0667、亮度_1.4667、YOLO_1、YOLO_3	0.9747	0.9722
原始色、亮度_1.0667、亮度_1.7333、YOLO_1、YOLO_3	0.9747	0.9722
原始色、亮度_1.0667、亮度_1.7333、YOLO_2、YOLO_3	0.9747	0.9722
原始色、亮度_0.9333、亮度_1.7333、YOLO_1、YOLO_3	0.9697	0.9722

表 19 排列組合特徵計數表

集成學習輸入特徵	Count
YOLO_3	15
原始色	10
YOLO_1	8
亮度_1.0667	6
亮度_1.7333	5
亮度_0.9333	5
熱區圖	3
YOLO_2	2
HSV	2
亮度_1.6	1
亮度_1.4667	1
亮度_1.8667	1
亮度_0.8	1

6.5 模型之績效與評估

本論文提出了多中藉由多種深度學習模型進行集成學習的方法，在將多種深度學習模型輸入至集成模型後，於表 20、21 訓練集及測試集混淆矩陣中發現經過集成以後可以有效的去修正單一模型所辨識的錯誤，於表 22、23 同樣也可以發現在訓練集及測試集績效當中，不只是準確度上有所提升，其餘的指標皆有提升。為了是研究更加有有效性，於最後再進行 5 折之交叉驗證證實此集成學習模型是有效的。

表 20 各別集成學習訓練集混淆矩陣

不同色彩空間			
Training population(N=198)		Real	
Predict	Positive	Positive	Negative
	Negative	97	6
不同亮度			
Training population(N=198)		Real	
Predict	Positive	Positive	Negative
	Negative	99	4
YOLO 集成			
Training population(N=198)		Real	
Predict	Positive	Positive	Negative
	Negative	94	9
C_5^{16} 組合			
Training population(N=198)		Real	
Predict	Positive	Positive	Negative
	Negative	101	2
	Positive	2	93
	Negative		

表 21 各別集成學習測試集混淆矩陣

不同色彩空間				
Testing population(N=36)		Real		
		Positive	Negative	
Predict	Positive	23	1	
	Negative	2	10	
不同亮度				
Testing population(N=36)		Real		
		Positive	Negative	
Predict	Positive	22	2	
	Negative	2	10	
YOLO 集成				
Testing population(N=36)		Real		
		Positive	Negative	
Predict	Positive	22	2	
	Negative	0	12	
C_5^{16} 組合				
Testing population(N=36)		Real		
		Positive	Negative	
Predict	Positive	23	1	
	Negative	0	12	

表 22 各模型訓練集實驗結果

Model	Accuracy	Precision	Recall	F1
C_5^{16} 組合	0.9798	0.9806	0.9806	0.9806
不同亮度	0.9697	0.9612	0.9802	0.9706
不同色彩空間	0.9293	0.9417	0.9238	0.9327
YOLO 集成	0.9091	0.9126	0.9126	0.9126

表 23 各模型測試集實驗結果

Model	Accuracy	Precision	Recall	F1
C_5^{16} 組合	0.9722	0.9583	1.0	0.9787
不同亮度	0.8889	0.9167	0.9167	0.9167
不同色彩空間	0.9167	0.9583	0.92	0.9388
YOLO 集成	0.9444	0.9126	1.0	0.9565

最後為了確保實驗嚴謹性，對各別集成學習模型進行交叉驗證，可以看見每一種模型於表 24 中各模型的平均準確度接可以大於 90%，因此可以證實本研究提出之方法能有效的辨識鼻咽部位影像。

表 24 集成學習交叉驗證實驗結果

模型	Acc.1	Acc.2	Acc.3	Acc.4	Acc.5	Mean
C_5^{16} 組合	0.9574	0.9787	0.9362	1.0	0.9565	0.9658
不同亮度	0.9787	0.9574	0.9149	0.9574	0.9348	0.9486
不同色彩空間	0.9362	0.9787	0.8511	0.9787	0.9787	0.9185
YOLO 集成	0.9149	0.9574	0.8723	0.9787	0.7609	0.8968

6.6 模型質化合理性說明

由於單一模型質化合理性已經於 6.3 小節說明，本小節僅介紹本研究所使用集成模型之質化合理性，將其分成真實資料為陽性被辨識成陽性、真實資料為陽性被辨識成陰性、真實資料為陰性被辨識成陽性、真實資料為陰性被辨識成陰性這幾個部分，以準確度最高之集成學習排列組合模型為例，並從表 25 觀察集成學習輸入端特徵與辨識結果及真實結果對應之關係，其中該表之特徵位說明的是假如顯示 11111，代表輸入端有 5 個陽性輸入結果至集成學習模型當中；假如顯示 10000，代表輸入端有 1 個陽性、4 個陰性結果輸入至集成學習模型當中。以下將說明其合理性原因。

表 25 正確辨識之集成學習輸入端特徵計數結果

辨識為陽性之特徵	Count	辨識為陰性之特徵	Count
11111	94	00000	63
11011	11	10000	16
10111	8	10001	4
11110	4	00001	4
10011	3	00011	3
11101	2	10010	2
10110	1	01001	2
11001	1	00100	2
		00010	2
		10100	2
		11100	1
		01010	1
		10101	1
		01011	1

首先為真實資料為陽性被辨識成陽性及真實資料為陰性被辨識成陰性的部分，從表 25 可以發現大部分輸入之特徵皆能符合多數決的法則，僅在真實資料為陰性被辨識成陰性中有少數不符合此條件。

接著以錯誤辨識的部分，真實資料為陽性被辨識成陰性、真實資料為陰性被辨識成陽性的部分，從表 26 顯示，其大多原因為原本輸入之特徵的模型就已經辨識錯誤，導致於輸入至集成學習模型後也接續判定錯誤。

表 26 錯誤辨識之集成學習輸入端特徵計數結果

辨識為陽性之特徵	Count	辨識為陰性之特徵	Count
11010	1	00000	1
10111	1	01011	1
11001	1	00011	1



第七章 結論與未來研究

由於鼻咽癌在前期較難以被發現，對於病患來說可能是一個極大的隱患，對於醫師長時間的診斷必然會造成生、心理的疲勞，因此為了使醫師及病患能夠提早的診斷出鼻咽癌病症，並提早規劃病患治療方式，一個好的醫療影像辨識是不可或缺的。在鼻咽部位影像辨識上，由於其影像雜訊及隨機性較高問題增加了辨識的難度，且過往的研究當中，過往的影像辨識方法並不能很有效的擷取影像之特徵，且大多的影像辨識方法僅依靠單一模型進行辨識，若是繼續以過往的方法進行辨識成效可能會有不彰的問題，因此本研究提出三個概念解決以上問題，首先藉由影像前處理的方法對雜訊去除，透過原始色、不同色彩空間、不同亮度三種影像找尋適合內視鏡環境，接著訓練多種 CNN 及 YOLO 模型並進行集成解決僅靠單一模型輸出的問題。最後本研究的實驗結果發現經過集成學習後，集成學習的績效遠比單一模型的績效要來的好，在集成學習的過程中也能有效的校正原本單一模型判斷錯誤的部分。

在實驗結果中，本論文證實提出的方法效能去修正單一模型所辨識之錯誤，並且在績效上遠比單一模型的績效要來的好很多，最終在集成學習模型當中訓練集準確度可以高達 0.9798，測試集準確度也可以高達 0.9722，因此能有效的辨識內視鏡鼻咽部為之影像。

在未來研究上，若是有足夠的樣本情況下，可以使用其他特徵工程的方式擷取影像之特徵，或者可以使用即時的方式去呈現，且若真的要使用 CNN 模型時建議在輸入模型前先使用一個影像篩選器，篩選器會篩選出品質較優或劣之影像，接著僅將品質較優之影像輸入 CNN 模型，即時的方式對於病患及醫生來說，醫生在現場藉由內視鏡檢查的同時可以透過模型即時的顯示患部的位置，能更快的去檢測是否有該病徵的發生，以便於醫生更好的去診斷病人。

參考文獻

- [1] 三軍總醫院耳鼻喉部(民 105 年 3 月 3 日)。**鼻咽癌**。民 109 年 11 月 5 日，取自：
<https://wwwv.tsgh.ndmctsgh.edu.tw/unit/10018/18172>
- [2] 台灣癌症基金會。**鼻咽癌**。民 109 年 11 月 7 日，取自：
<https://www.canceraway.org.tw/page.asp?IDno=2225>
- [3] Abd Ghani, M. K., Mohammed, M. A., Arunkumar, N., Mostafa, S. A., Ibrahim, D. A., Abdullah, M. K., ... & Burhanuddin, M. A. (2020). Decision-level fusion scheme for nasopharyngeal carcinoma identification using machine learning techniques. *Neural Computing and Applications*, 32(3), 625-638.
- [4] Adjed, F., Faye, I., Ababsa, F., Gardezi, S. J., & Dass, S. C. (2016, November). Classification of skin cancer images using local binary pattern and SVM classifier. In *AIP Conference Proceedings* (Vol. 1787, No. 1, p. 080006). AIP Publishing LLC.
- [5] Akbari, M., Mohrekesh, M., Rafiei, S., Soroushmehr, S. R., Karimi, N., Samavi, S., & Najarian, K. (2018, July). Classification of informative frames in colonoscopy videos using convolutional neural networks with binarized weights. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 65-68). IEEE.
- [6] Alfed, N., & Khelifi, F. (2017). Bagged textural and color features for melanoma skin cancer detection in dermoscopic and standard images. *Expert Systems with Applications*, 90, 101-110.
- [7] Al-Masni, M. A., Al-Antari, M. A., Park, J. M., Gi, G., Kim, T. Y., Rivera, P., ... & Kim, T. S. (2018). Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system. *Computer methods and programs in biomedicine*, 157, 85-94.
- [8] Amin, J., Sharif, M., Raza, M., Saba, T., & Anjum, M. A. (2019). Brain tumor detection

- using statistical and machine learning method. *Computer methods and programs in biomedicine*, 177, 69-79.
- [9] Amit, G., Ben-Ari, R., Hadad, O., Monovich, E., Granot, N., & Hashoul, S. (2017, March). Classification of breast MRI lesions using small-size training sets: comparison of deep learning approaches. In *Medical Imaging 2017: Computer-Aided Diagnosis* (Vol. 10134, p. 101341H). International Society for Optics and Photonics.
- [10] Amulya, C., & Prathibha, G. (2016). MRI Brain Tumor Classification Using SURF and SIFT Features. *International Journal for Modern Trends in Science and Technology*, 2(7), 123-127.
- [11] Bakheet, S. (2017). An SVM framework for malignant melanoma detection based on optimized HOG features. *Computation*, 5(1), 4.
- [12] Bardou, D., Zhang, K., & Ahmad, S. M. (2018). Classification of breast cancer based on histology images using convolutional neural networks. *IEEE Access*, 6, 24680-24693.
- [13] Cao, Z., Duan, L., Yang, G., Yue, T., & Chen, Q. (2019). An experimental study on breast lesion detection and classification from ultrasound images using deep learning architectures. *BMC medical imaging*, 19(1), 51.
- [14] Cheng, J., Huang, W., Cao, S., Yang, R., Yang, W., Yun, Z., ... & Feng, Q. (2015). Enhanced performance of brain tumor classification via tumor region augmentation and partition. *PloS one*, 10(10), e0140381.
- [15] Dalal, N. (2006). *Finding people in images and videos* (Doctoral dissertation).
- [16] Daoud, B., Morooka, K. I., Kurazume, R., Leila, F., Mnejja, W., & Daoud, J. (2019). 3D segmentation of nasopharyngeal carcinoma from CT images using cascade deep learning. *Computerized Medical Imaging and Graphics*, 77, 101644.
- [17] David, E., Boia, R., Malaescu, A., & Carnu, M. (2013, July). Automatic colon polyp detection in endoscopic capsule images. In *International Symposium on Signals, Circuits and Systems ISSCS2013* (pp. 1-4). IEEE.

- [18] De Sampaio, W. B., Silva, A. C., de Paiva, A. C., & Gattass, M. (2015). Detection of masses in mammograms with adaption to breast density using genetic algorithm, phylogenetic trees, LBP and SVM. *Expert Systems with Applications*, 42(22), 8911-8928.
- [19] Deepika, K., Bodapati, J. D., & Srihitha, R. K. (2019). An Efficient Automatic Brain Tumor Classification Using LBP Features and SVM-Based Classifier. In *Proceedings of International Conference on Computational Intelligence and Data Engineering* (pp. 163-170). Springer, Singapore.
- [20] Désir, C., Petitjean, C., Heutte, L., Thiberville, L., & Salaün, M. (2012). An SVM-based distal lung image classification using texture descriptors. *Computerized Medical Imaging and Graphics*, 36(4), 264-270.
- [21] George, J., Skaria, S., & Varun, V. V. (2018, February). Using YOLO based deep learning network for real time detection and localization of lung nodules from low dose CT scans. In *Medical Imaging 2018: Computer-Aided Diagnosis* (Vol. 10575, p. 105751I). International Society for Optics and Photonics.
- [22] Gheisari, S., Catchpoole, D. R., Charlton, A., Melegh, Z., Gradhand, E., & Kennedy, P. J. (2018). Computer aided classification of neuroblastoma histological images using scale invariant feature transform with feature encoding. *Diagnostics*, 8(3), 56.
- [23] Ghosh, S., Bandyopadhyay, A., Sahay, S., Ghosh, R., Kundu, I., & Santosh, K. C. (2021). Colorectal Histology Tumor Detection Using Ensemble Deep Neural Network. *Engineering Applications of Artificial Intelligence*, 100, 104202.
- [24] Gueye, L., Yildirim-Yayilgan, S., Cheikh, F. A., & Balasingham, I. (2015, September). Automatic detection of colonoscopic anomalies using capsule endoscopy. In *2015 IEEE International Conference on Image Processing (ICIP)* (pp. 1061-1064). IEEE.
- [25] Guo, X., Zhang, N., Guo, J., Zhang, H., Hao, Y., & Hang, J. (2019). Automated polyp segmentation for colonoscopy images: A method based on convolutional neural networks and ensemble learning. *Medical physics*, 46(12), 5666-5676.
- [26] Halicek, M., Lu, G., Little, J. V., Wang, X., Patel, M., Griffith, C. C., ... & Fei, B. (2017). Deep convolutional neural networks for classifying head and neck cancer using

- hyperspectral imaging. *Journal of biomedical optics*, 22(6), 060503.
- [27] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
- [28] Iwahori, Y., Hattori, A., Adachi, Y., Bhuyan, M. K., Woodham, R. J., & Kasugai, K. (2015). Automatic detection of polyp using hessian filter and hog features. *Procedia computer science*, 60, 730-739.
- [29] Jha, D., Ali, S., Tomar, N. K., Johansen, H. D., Johansen, D., Rittscher, J., ... & Halvorsen, P. (2021). Real-time polyp detection, localization and segmentation in colonoscopy using deep learning. *IEEE Access*, 9, 40496-40510.
- [30] Kaplan, K., Kaya, Y., Kuncan, M., & Ertunç, H. M. (2020). Brain tumor classification using modified local binary patterns (LBP) feature extraction methods. *Medical Hypotheses*, 109696.
- [31] Kavitha, J. C., Suruliandi, A., Nagarajan, D., & Nadu, T. (2017). Melanoma detection in dermoscopic images using global and local feature extraction. *Int J Multimed Ubiquit Eng*, 12(5), 19-28.
- [32] Khan, M. A., Rubab, S., Kashif, A., Sharif, M. I., Muhammad, N., Shah, J. H., ... & Satapathy, S. C. (2020). Lungs cancer classification from CT images: An integrated design of contrast based classical features fusion and selection. *Pattern Recognition Letters*, 129, 77-85.
- [33] King, A. D., Vlantis, A. C., Bhatia, K. S., Zee, B. C., Woo, J. K., Tse, G. M., ... & Ahuja, A. T. (2011). Primary nasopharyngeal carcinoma: diagnostic accuracy of MR imaging versus that of endoscopy and endoscopic biopsy. *Radiology*, 258(2), 531-537.
- [34] Korkmaz, S. A., & Binol, H. (2018). Classification of molecular structure images by using ANN, RF, LBP, HOG, and size reduction methods for early stomach cancer detection. *Journal of Molecular Structure*, 1156, 255-263.

- [35] Kumar, R., Srivastava, R., & Srivastava, S. (2015). Detection and classification of cancer from microscopic biopsy images using clinically significant and biologically interpretable features. *Journal of medical engineering*, 2015.
- [36] Lee, J. Y., Jeong, J., Song, E. M., Ha, C., Lee, H. J., Koo, J. E., ... & Byeon, J. S. (2020). Real-time detection of colon polyps during colonoscopy using deep learning: systematic validation with four independent datasets. *Scientific Reports*, 10(1), 1-9.
- [37] Li, B., & Meng, M. Q. H. (2012). Automatic polyp detection for wireless capsule endoscopy images. *Expert Systems with Applications*, 39(12), 10952-10958.
- [38] Liew, W. S., Tang, T. B., Lin, C. H., & Lu, C. K. (2021). Automatic colonic polyp detection using integration of modified deep residual convolutional neural network and ensemble learning approaches. *Computer Methods and Programs in Biomedicine*, 206, 106114.
- [39] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2117-2125).
- [40] Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2), 91-110.
- [41] Mishra, V., Singh, Y., & Rath, S. K. (2019, March). Breast Cancer detection from Thermograms Using Feature Extraction and Machine Learning Techniques. In *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)* (pp. 1-5). IEEE.
- [42] Mohammed, M. A., Abd Ghani, M. K., Hamed, R. I., & Ibrahim, D. A. (2017). Review on Nasopharyngeal Carcinoma: Concepts, methods of analysis, segmentation, classification, prediction and impact: A review of the research literature. *Journal of Computational Science*, 21, 283-298.
- [43] Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, 24(7), 971-987.

- [44] Pogorelov, K., Ostroukhova, O., Jeppsson, M., Espeland, H., Griwodz, C., de Lange, T., ... & Halvorsen, P. (2018, June). Deep learning and hand-crafted feature based approaches for polyp detection in medical videos. In *2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS)* (pp. 381-386). IEEE.
- [45] Pogorelov, K., Riegler, M., Eskeland, S. L., de Lange, T., Johansen, D., Griwodz, C., ... & Halvorsen, P. (2017). Efficient disease detection in gastrointestinal videos—global features versus neural networks. *Multimedia Tools and Applications*, 76(21), 22493-22525.
- [46] Prasath, V. B. (2017). Polyp detection and segmentation from video capsule endoscopy: A review. *Journal of Imaging*, 3(1), 1.
- [47] Redmon, J., & Farhadi, A. (2017). YOLO9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7263-7271).
- [48] Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. *Computer Vision and Pattern Recognition (CVPR)*, 126-134.
- [49] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788).
- [50] Safdar, M. F., Alkobaisi, S. S., & Zahra, F. T. (2020). A Comparative Analysis of Data Augmentation Approaches for Magnetic Resonance Imaging (MRI) Scan Images of Brain Tumor. *Acta Informatica Medica*, 28(1), 29.
- [51] Sánchez-Peralta, L. F., Bote-Curiel, L., Picón, A., Sánchez-Margallo, F. M., & Pagador, J. B. (2020). Deep learning to find colorectal polyps in colonoscopy: A systematic literature review. *Artificial Intelligence In Medicine*, 101923.
- [52] Savitha, G., & Jidesh, P. (2019). A fully-automated system for identification and classification of subsolid nodules in lung computed tomographic scans. *Biomedical Signal Processing and Control*, 53, 101586.
- [53] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017).

- Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision* (pp. 618-626).
- [54] Shiji, T. P., Remya, S., & Thomas, V. (2017). Computer aided segmentation of breast ultrasound images using scale invariant feature transform (SIFT) and bag of features. *Procedia computer science*, 115, 518-525.
- [55] Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1), 1-48.
- [56] Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2017, February). Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 31, No. 1).
- [57] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1-9).
- [58] Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., & Liang, J. (2017). On the necessity of fine-tuned convolutional neural networks for medical imaging. In *Deep Learning and Convolutional Neural Networks for Medical Image Computing* (pp. 181-193).
- [59] Telea, A. (2004). An image inpainting technique based on the fast marching method. *Journal of graphics tools*, 9(1), 23-34.
- [60] Urban, G., Tripathi, P., Alkayali, T., Mittal, M., Jalali, F., Karnes, W., & Baldi, P. (2018). Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy. *Gastroenterology*, 155(4), 1069-1078.
- [61] Vijayarajeswari, R., Parthasarathy, P., Vivekanandan, S., & Basha, A. A. (2019). Classification of mammogram for early detection of breast cancer using SVM classifier and Hough transform. *Measurement*, 146, 800-805.
- [62] Yu, Y., Ma, J., Zhao, W., Li, Z., & Ding, S. (2020). MSCI: A multistate dataset for

colposcopy image classification of cervical cancer screening. *International Journal of Medical Informatics*, 104352.

- [63] Yuan, Y., & Meng, M. Q. H. (2014, September). A novel feature for polyp detection in wireless capsule endoscopy images. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 5010-5015). IEEE.
- [64] Yuan, Y., Li, B., & Meng, M. Q. H. (2015). Improved bag of feature for automatic polyp detection in wireless capsule endoscopy images. *IEEE Transactions on automation science and engineering*, 13(2), 529-535.
- [65] Yuan, Y., Li, B., & Meng, M. Q. H. (2016). WCE abnormality detection based on saliency and adaptive locality-constrained linear coding. *IEEE Transactions on Automation Science and Engineering*, 14(1), 149-159.

