

醫學影像處理期末專題_骨齡辨識深度學習

學生：F94101070 陳妍妏、F94106088 黃鈺雯

一、摘要

近年來，越來越多父母在意孩子的身高，擔心這可能對未來的發展產生影響。而為了評估兒童的骨骼發育狀況，兒童內分泌科醫師通常使用骨齡作為主要的診斷工具。然而，傳統的評估方法存在著主觀判斷的差異，需要依賴醫師的經驗才能做出準確的判斷。

因此，許多研究紛紛投入開發利用深度學習辨識骨齡的系統，以提高評估的客觀性和精確性。而本專案主要是希望透過文獻回顧，結合不同研究所使用的方法，並相互比較以找出最合適的模型架構，完成模型訓練後，我們將進一步設計使用者介面，將評估結果視覺化呈現。不僅有助於提供醫師更直觀、客觀的工具，協助他們全面評估兒童的骨齡發育情況，也能給予家長科學的參考，使其更有效地應對對孩子身高的關切，促進健康的成長與發展。

二、研究動機與研究問題

(一) 現階段臨床問題

現階段臨床評估方法主要分為 GP 法與 TW 法[1]，如下：

1. GP 法 (Greulich and Pyle method)

主要評估方法是將受試者左手的 X 光對照標準化圖譜，來回比對，找出最類似的骨齡圖，來判定孩子的骨齡。但該圖譜資料久未更新，可能與現代人骨骼發展特徵有所出入，更需考慮資料集與各地人種骨骼發展之差異將影響評估結果。

2. TW 法 (Tanner and Whitehouse)[2]

主要評估方法是將左手 X-ray 圖像拆分成幾個特徵判斷區域進行評估計分，最後再依加總分數進行判斷，雖較 GP 法更準確，但評估上更花時間，因此較少被使用。

綜合上述，可觀察到傳統評估方式的準確度會因不同醫師的主觀判斷而有差異，且評估極為耗時，因此自動化骨齡檢測系統對於臨床有一定的需求。

(二) 深度學習應用未滿足需求

深度學習的技術可透過自動化作業消除臨床上診斷骨齡的主觀誤差，目前研究精準度已接近人工辨識的評估結果，說明深度學習在實現骨齡檢測的自動化有極大的潛力[3]。然而，儘管目前有許多單位紛紛投入深度學習檢測骨齡的領域中，深度學習檢測骨齡仍未被廣泛應

用在臨床，僅有少數醫療院所引入相關系統協助看診，代表可能還存在著某些疑慮。而我們查閱資料發現許多文獻使用的方法都值得參考，但沒有人去整合，我們希望在期末專題中嘗試不同文獻的方法，整合出最合適的骨齡辨識系統。

三、文獻回顧與探討

(一) 傳統臨床骨齡判斷方法

骨齡評估是對骨骼的發育特徵進行定量的評估計算，以量化骨骼發育程度。傳統上由醫師進行判斷，通常藉由手部的 X-ray 圖像作為評估的依據，評估對象通常為小兒，5~19 歲是手部骨骼發育差異較大的年齡區間，也常藉由骨齡評估判斷青春期的延遲、身材矮小或性早熟問題。常見的判斷方法為 GP 法與 TW 法。

(二) 深度學習用於骨齡的評估

早期的研究由於醫學影像的資料集較難蒐集，因此會考慮到使用遷移學習作為訓練方法。2017 年，Lee 等[4]研發了一套包含興趣區域分離、圖片標準化及預處理、自動化骨齡辨識及一鍵產生結構化 X-ray 報告的 DL-BAA 系統，最後利用 GoogLeNet 作為框架進行微調學習。其在測試集影像中的女性準確率為 57.32%，男性為 61.40%，1 歲誤差範圍內的準確率為 90.39%，2 歲為 98.11%；女性均方根誤差(RMSE)為 0.93 歲，男性為 0.82 歲。且作者還使用輸入遮蔽的方法繪製熱圖，呈現訓練模型用於執行 BAA 的特徵，研究該系統對手部骨骼的哪些區域較敏感。此 DL-BAA 系統雖極大提高了骨齡分析的速度，但該研究訓練集中缺乏 0~4 歲兒童及手部骨骼畸形群體的影像學圖像，限制了該系統的適用範圍。

2017 年，Zhou 等[5]使用定義各種感興趣區域(ROI)和基於遷移學習的深度卷積神經網路(DCNN)進行骨齡分類。研究使用洛杉磯兒童醫院創建的數位手圖譜作為資料集，該資料集中包含 1390 張 0~18 歲兒童的左手 X 光片，作者首先對每個自定義出的 ROI 進行基於直方圖的對比均衡、旋轉、翻轉等增強預處理。再使用基於 ImageNet 預先訓練的 DCNN，利用遷移學習對每個 ROI 區塊進行微調，並為每個 ROI 建立一個分類模型，最後採用模型融合方法得到骨齡分類的最終模型。這項研究測試資料後達到最低的 MAE 值為 0.72 歲，雖然資料集較小，但預處理做的好，使得結果不至於落差太大。

2018 年，Iglovikov 等[6]透過 U-Net 模型分割手部區域來去除背景，並進行標準化對比度。另外，透過檢測關鍵點，應用仿射變換將影像轉換到一個共同的座標空間，此作法不僅提高數據品質，還讓識別手部不同區域的準確度有所提升。除此之外，他們使用手部影像的

不同部分訓練多個深度網絡，以評估不同的手部骨骼如何在四個主要骨骼發育階段對模型的性能做出貢獻。為了評估預處理對預測性能的影響，他們在原始影像、分割和歸一化影像以及分割、歸一化和配準影像上評估回歸網路，相應的 MAE 為 31.56、8.76 和 8.08 個月，相應地證明瞭預處理帶來的性能改進。

2020 年，Ari 和 Petrus[7]提出了一個基於區域的特徵連接層(RB-FCL)，運用基於區域的捲積神經網路(R-CNN)對手部 X 光進行自動分割，將手部 X 線骨骼區域分割為尺桡骨、腕骨、掌指骨、指骨和骨骺 5 個區域，使用 DL 模型對手部 X 射線骨骼各區域特徵提取，創建了一個 DL 模型，產生具有 1 024 個密集特徵的 RB-FCL，取得更具代表性的區域模型特徵，這些 FCL 是手部 X 線分割區域的重要輸出，使迴歸模型能更好地預測骨齡，產生的最小 MAE 為 6.97 個月。

綜合上述，可以瞭解到資料的標準化、影像前處理與 ROI 分割區域有助於提高模型訓練的準確度，而我們著重參考了 U-Net 手部分割的方法，並輔以課程中影像處理的技術，作為主要的前處理。

四、研究方法及步驟

(一)資料集

本專案使用的資料集是來自北美放射學會(RSNA)挑戰的公共訓練資料集[8]。該資料集包含 12,611 張 X-ray 照片，以及涵蓋每張 X-ray 照片對應骨齡和性別資訊的 csv 檔案。其中，骨齡年齡分佈比例如圖 1 所示。

此外，資料集中 X-ray 照片尺寸、對比、翻轉等參數皆不同，因此在進行深度學習訓練時，需進行影像前處理。

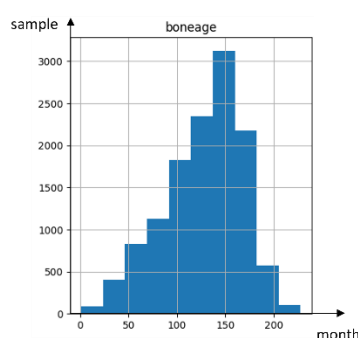


圖 1. RSNA 骨齡分佈圖

(二)資料前處理

如圖 1，可以發現資料集中在兩側的資料量偏少，通常在深度學習分析會將兩側做數據增強(Data augmentation)，以達到資料的平均性。然而，因骨齡分析對於影像原始的長寬比十分注重，而無法使用調整

圖片參數來擴充資料的方法，因此，資料前處理我們只進行將在原先資料集中的 csv 加入一個圖片路徑的 column，以方便後續訓練，如圖 2 所示。

	id	boneage	male	Image Path
0	1377	180	False	1377.png
1	1378	12	False	1378.png
2	1379	94	False	1379.png
3	1380	120	True	1380.png
4	1381	82	False	1381.png

圖 2. 資料前處理

(三)影像預處理

本專案的影像處理的總流程如圖 3，每個步驟的生成如圖 4：

1. 手部分割

因資料集本身的影像有背景的干擾，為了讓影像更具焦在手部的分析，我們先將資料集標準化後，參考[7]的方式，使用預訓練的 U-Net 模型[9]來偵測每張 X-ray 影像中的手部邊界，以協助我們生成每張影像的遮罩，生成後再將遮罩套入資料集的影像中，達到手部分割的效果。

2. 標準化

在[7]的文獻中提到，使用 CNN 進行關鍵點偵測，可有效標準化所有影像，是解決手部位置不一的好方法，但礙於時間壓力，本專案用以下方式取代之。經預訓練的 U-Net 模型遮罩覆蓋後，我們裁剪出手部區域，將其等比例縮放後，置放於 224x224 的黑色圖片中央，以避免長寬比被調整，以致資料失真。

3. 特徵提取

為使骨骼更為清楚，本專案首先調整所有影像的灰階值，統一影像的平均強度。再使用 CLACHE 對比增強，將骨頭顏色區別出來。最後經降噪後，使用銳利化濾波，更為凸顯骨骼的邊界，用以作為模型訓練前的特徵提取。

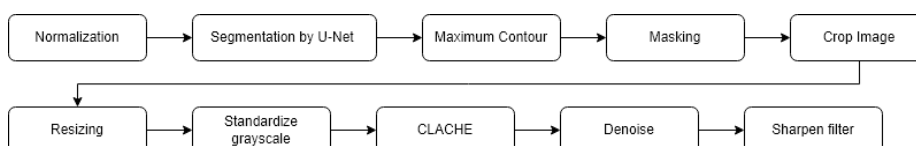


圖 3. 影像預處理流程圖

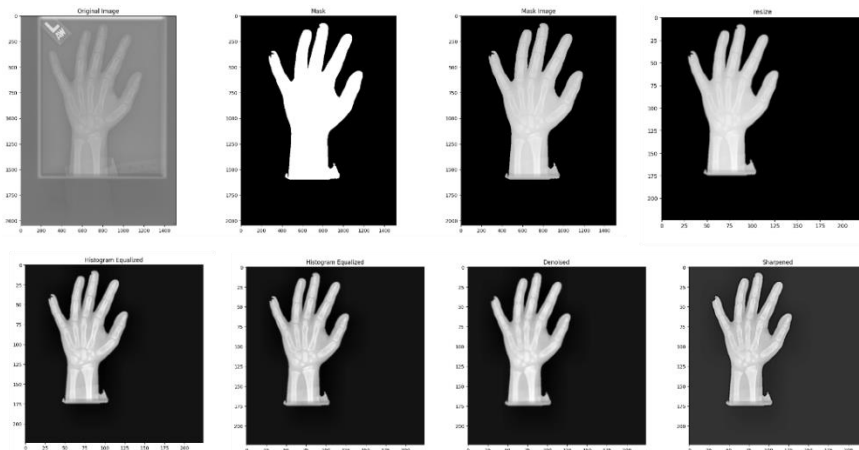


圖 4. 影像預處理結果圖

(四)模型訓練

經文獻回顧後發現大部分的骨齡分析皆是利用 CNN 模型進行實作，因此本專案也嘗試透過 CNN 進行訓練與測試。本架構使用 Adam 優化器，損失函數選用均方誤差(MSE)，並使用平均絕對誤差(MAE)來當作是效能評估工具。

模型架構如圖 5 所示，而模型架構輸入為 $224 \times 224 \times 1$ ，經過四層捲積層與池化層，各層皆使用 ReLU 作為 Activation function，最後輸出經由 flatten 後，再經過 3 層隱藏層，隱藏層間使用 Dropout 避免資料 overfitting。

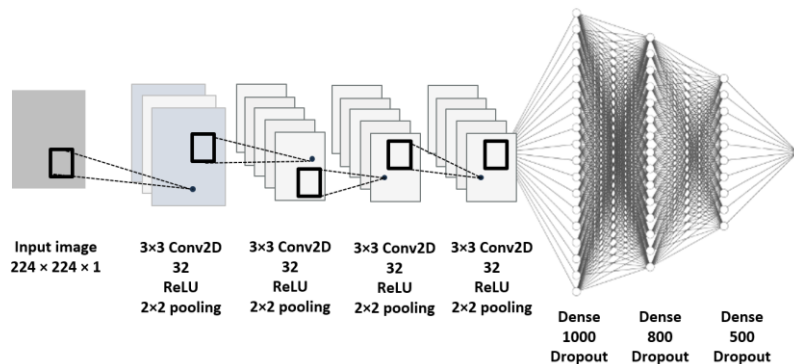


圖 5. CNN 模型架構圖

(五)使用者介面

本專案使用基於 Flask 框架的 Web 應用程式，接收到使用者上傳的影像，會送到後端進行上述提到的影像預處理，再套入預先訓練好的深度學習模型進行骨齡預測，最後將處理後的影像及預測結果送回前端，而介面主要由 HTML 與 CSS 進行設計。

五、研究結果

(一)CNN 辨識模型訓練結果

模型訓練的結果如圖 6 所示，圖中顯示本專案使用的 120 個 epochs 期間損失 (Loss) 和平均絕對誤差 (MAE) 的變化情況。訓練過程表現出 Loss 和 MAE 有收斂的趨勢，這表明模型在這個訓練過程中逐漸學習到了數據的特徵，並在最後的 epochs 中取得了相對穩定的結果。

而最終的 MAE 結果 16.4681 月，雖與文獻回顧中提到的最低 MAE 為 6.97 有明顯的差距，但還是可接受的訓練結果。

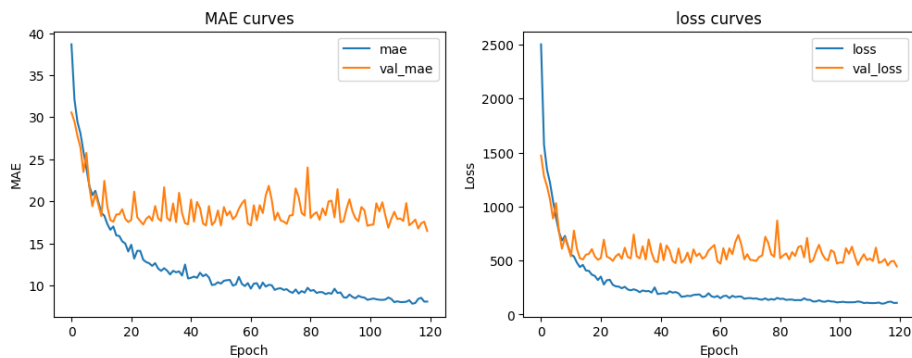


圖 6. CNN 模型預測結果評估

(二)使用者介面設計

圖 7 為最終設計出的網頁，使用者可在首頁中點選「點我去測試」，爾後會導向骨齡測試的頁面，當使用者匯入手部 x-ray 影像並輸入實際年齡，網頁會輸出評估的骨齡值，與使用者實際年齡的差值，並生成預處理後的影像。為了讓使用者操作較為容易，網頁設計相較簡單。

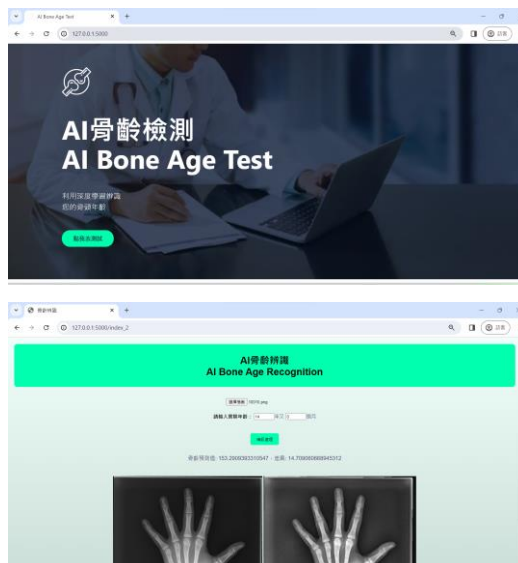


圖 7. 使用者介面展示

六、 討論

(一)訓練結果討論

1. 我們訓練的 MAE 結果為 16.4681，與論文提及的 6.97 來的大，推論原因是一開始影像預處理手部定位的問題。[7]文獻中提到，將手部方向、大小統一化，可以使精確度變高，是我們未來時間允許下，可以努力的方向。
2. 資料集影像經過前處理所訓練出的模型，MAE 比使用未經前處理影像所訓練的模型低，說明影像經過適當的前處理有可能提升模型準確度。
3. 我們將訓練完成的模型輸出後進行測試，發現骨齡較小的測試結果時常差距較多，推測是因為低年齡段的影像資料較少。針對此問題有論文提及資料擴充的方式來解決，但未有比較資料擴充是否會對模型訓練結果造成正向影響，因此我們未進行嘗試。
4. 除了適當的前處理、充足的訓練資料外，也有文獻提及調整模型判讀影像的 ROI 助於提升模型準確度，形同專業醫師進行診斷的方式。此方法可將注意力集中在骨骼發展型態的差異，甚至是常用來進行判斷的區域。相比之下，直接用 CNN 進行分類可能受到手部姿勢、影像位置、軟組織型態等區域影響，或是因標準化不夠完善而影響模型判斷，產生誤差，因此我們推斷調整模型判讀影像的 ROI 對模型的可靠性與穩定性能夠大大提升。

(二)使用者介面討論

目前尚無免費的骨齡檢測網頁公開於瀏覽器中，我們期望未來重新修正前處理方式與模型架構，待準確度更為提升時，上架於網頁上，讓關心孩子健康的父母都能免費享有此資源。

七、 結論

在本專案中，我們在影像預處理中運用影像處理的技術，得到特徵提取後的資料集；在自動化檢測骨齡處使用深度學習模型，得到相對的預測值；最後在使用者介面使用 HTML、CSS 等前端技術，將結果透過網頁呈現出來。雖然目前礙於時間不足，無法進行複雜但相對好的前處理，也沒有實際透過整合不同論文的模型找到最低 MSE，但未來時間允許下，我們會繼續完成，並上架於公開平台，免費提供給大眾使用。

八、 附件

GitHub 連結：<https://github.com/yanwen1107/BoneAgeProject>

九、參考文獻

1. 黃世綱. (2021, August 12). 骨齡的故事 - 從骨齡圖譜的前世與今生談骨齡判讀. <https://www.drgrowup.tw/2021/08/bone-age-assessment-history.html>
2. 徐永釗. (2021, August 12). 詳解骨齡檢測 G-P 骨齡圖譜 (男生版). 知乎. <https://zhuanlan.zhihu.com/p/561345247>
3. 吳莉婷, 吳瑞, 李書香, 等. 基於深度學習的人工智能技術與骨齡評估的研究現狀 [J]. 中華實用兒科臨床雜誌, 2023, 38(5) : 394-397. DOI: 10.3760/cma.j.cn101070-20211213-01456.
4. Lee, H., Tajmir, S., Lee, J. et al. Fully Automated Deep Learning System for Bone Age Assessment. J Digit Imaging 30, 427–441 (2017). <https://doi.org/10.1007/s10278-017-9955-8>
5. J. Zhou, Z. Li, W. Zhi, B. Liang, D. Moses and L. Dawes, "Using Convolutional Neural Networks and Transfer Learning for Bone Age Classification," 2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA), Sydney, NSW, Australia, 2017, pp. 1-6, doi: 10.1109/DICTA.2017.8227503.
6. Iglovikov, V.I., Rakhlin, A., Kalinin, A.A., Shvets, A.A. (2018). Paediatric Bone Age Assessment Using Deep Convolutional Neural Networks. In: Stoyanov, D., et al. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. DLMIA ML-CDS 2018 2018. Lecture Notes in Computer Science(), vol 11045. Springer, Cham. https://doi.org/10.1007/978-3-030-00889-5_34
7. Wibisono, A., Mursanto, P. Multi Region-Based Feature Connected Layer (RB-FCL) of deep learning models for bone age assessment. J Big Data 7, 67 (2020). <https://doi.org/10.1186/s40537-020-00347-0>
8. Dataset: <https://www.kaggle.com/datasets/kmader/rsna-bone-age>
9. Hand mask for the RSNA bone age dataset <https://zenodo.org/records/7611677>