

東海大學資訊工程學系

專題報告書

利用深度學習開發骨質疏鬆檢測系統

Implement an Osteoporosis classification system with using  
X-ray radiography via Deep learning

指導教授：

陳仕偉

專題學生：

陳俊兆 s09350711

吳易儒 s09350220

中華民國

112 年

12 月

16 日

## 中文摘要

本次實驗以深度學習與骨質疏鬆為主軸，並嘗試以深度學習的方式來解決骨質疏鬆中隱藏的諸多問題，如成本、空間、速度和普及率等等，希望能從根本解決骨質疏鬆對社會帶來的問題，並同時創造經濟價值。

骨質疏鬆症(Osteoporosis)是一種常見的慢性疾病，主要是因為年紀增長的骨質流失所造成的，這些微結構的破壞，使得年長者的生命受到強烈的威脅，其中又以飲食不均、缺乏規律運動、或有家族遺傳相關史的人群更加嚴重。

骨質疏鬆在臨床上的診斷和監控，目前以雙能量 X 光吸收儀(DEXA)所測得之骨礦物質密度(BMD)為黃金標準。此次實驗欲利用 X 光影像較高解析度之優勢，並結合真實臨床數據，通過深度學習訓練一個骨質疏鬆檢測系統，並建立一個使用者介面，為骨質疏鬆進行簡單且有效的分析，幫助病患即早發現並接受相關治療。深度學習不但提高了骨質疏鬆早期判斷的可行性，也有助於制定更準確與完善的醫療方向，同時解決了以往必須透過精密醫療儀器才能判斷骨質疏鬆的問題，以最大程度的漸少骨質疏鬆對病患帶來的生活不便與威脅。

此篇報告書將詳細介紹此次研究目的和其相關知識與模型選擇的原因與應用並同時研究其市場需求。願能更了解應對骨質疏鬆的方式，為病患提供更好的健康保障。

**關鍵詞：**Osteoporosis、DEXA、BMD、深度學習、Django、Image Classification

## 致謝

本畢業論文之改善與完成，此篇首先要先感謝我們的指導老師陳仕偉教授對我們專題的細心教導與不斷的鼓勵，您的春風化雨與盡心盡力我們都能感受其中，半年多來，從題目訂定到文獻研究，再到研究方向的選擇，觀念架構之建立等等，吾師以細心的態度讓我們受益良多，才能使我們如期完成此次艱難的任務，高山不移，碧水長流，我師恩澤，將在心永留，故特以此段落來對恩師以文字上表達我們銘感五內之心情。

同時也要感謝黃育仁教授在大三下課程影像視訊處理應用中對我們專題的意見說明與方向指導，老師用心提出的方法與改善方式，我們也都實際應用於專題的開發之中，並也有以此獲得更好的成效與多方面的改善，即使不是您的專題學生，老師依然鞠躬盡瘁且教導有方，謹在此致以最深的謝意。

另外也要特別感謝東海大學資訊工程學系對本研究的資金贊助、台中榮民總醫院的資料彙整與圖像作為本實驗的基礎、采風智匯股份有限公司的技術支援和相關知識的討論與幫助，特此一併致謝。

當然也必須感謝組員之間的戮力合作與共同扶持，在製作專題期間儘管遇到許多困難，我們組員之間仍保持良好的溝通與相互鼓勵，尤其在撰寫論文的最後階段，需要大量資料交流與意見交換的同時仍保持正向的態度，衷心感謝組員間用心的付出。

# 目錄

中文摘要	i
致謝	ii
目錄	iii
圖目錄	v
表目錄	vi
第壹章 緒論：	1
第一節 研究背景	1
第二節 研究動機	1
第三節 研究目的	2
第四節 研究範圍	2
第五節 研究方法概述	2
第貳章 相關文獻回顧：	3
第一節 研究相關論文	3
第二節 架構相關論文	3
第三節 方法相關論文	4
第參章 研究流程擬定：	5
第一節 研究方法	5
第二節 方法架構	6
第肆章 實驗過程論述：	9
第一節 實驗流程	9
第二節 實驗結果	13
第伍章 結論：	14
第一節 結論	14
第二節 研究限制討論	14
第三節 研究成果	15
第四節 研究貢獻	15

第五節 未來展望與後續研究 .....	16
參考文獻 .....	18

## 圖目錄

圖 3-1 流程圖 .....	5
圖 3-2 標註部位 .....	7
圖 3-3 金屬支架 .....	7
圖 3-4 歪斜情況 .....	7
圖 3-5 Django 架構圖 .....	8
圖 4-1 效能圖 .....	9
圖 4-2 折線圖 .....	9
圖 4-3 效能圖 .....	10
圖 4-4 折線圖 .....	10
圖 4-5 Loss 圖表 .....	11
圖 4-6 效能圖 .....	13
圖 4-7 折線圖 .....	13
圖 4-8 結果效能圖 .....	13
圖 5-1 網站主頁 .....	16
圖 5-2 使用者手冊介面 .....	16
圖 5-3 偵測到骨質疏鬆並附上自我評量表 .....	17
圖 5-4 網絡掛號 .....	17

## 表目錄

表格 4.1 各模型對於無骨質疏鬆的測試集數據.....	11
表格 4.2 各模型對於有骨質疏鬆的測試集數據.....	11

## 第壹章 緒論：

### 第一節 研究背景

AI 迅速發展的年代中，有越來越多能夠由 AI 解決的問題被挖掘，而在這塊稱為人工智慧的寶藏中，醫療 AI 便是其中之一，醫療 AI 不僅僅能提升全世界人民的生活品質，也能創造相當高的利益和經濟發展，同時跳出昂貴傳統醫療的窠臼，進而實現人人皆能享有生命與醫療保障的願景。

我們著手的目標是發現生活中不常被眾人關注，但卻會在暗中危急生命且醫療金額高昂的痼疾，於是我們選擇觀察「骨質疏鬆」會帶來的潛在危害和如何利用 AI 盡早發覺，來預防骨鬆意外發生。

### 第二節 研究動機

我們以以下兩點來說明我們的研究動機：

1. 在骨質疏鬆的領域中，骨質礦物密度值（BMD）是現今被認定的黃金標準數值[1]，而唯一被 WHO 推薦為 BMD 檢測標準的機器「雙能 X 射線吸收骨密度儀（DEXA）」造價卻十分昂貴[2]，並不是所有診所甚至地區性醫院能夠擁有的醫療設備，即使是擁有 DEXA 骨密儀的醫院，也是長期處於供不應求的狀況，我們的第一個動機便是為了嘗試解決這種長期供需不平衡的現況，並且運用 AI 來創造相對較低的價格來滿足市場需求。

2. 「根據國際骨質疏鬆基金會 2022 調查顯示：全世界 50 歲以上的人群中，女性有 1/3 會出現骨質疏鬆性骨折，男性有 1/5 會出現骨質疏鬆性骨折」[3]，其中又以 65 歲以上佔大多數，而這驚愕的人數比例竟只包含已發生骨折的人數，有更是還沒有發生意外的骨鬆潛在族群，這些骨鬆黑數面臨的不啻是一種威脅生命安全的疾病，甚至有可能沒有發覺身體其他疾病的警訊（如 糖尿病、腎病、肝病、高血鈣、甲狀腺機能亢進、風溼性關節炎、僵直性脊椎炎及骨癌 等都有可能併發骨質疏鬆的狀況），為了解決這個問題並即早發現治療，我們期望用深度學習的方式來創造一種能夠「降低負擔價格」與「提升普及率」的方法來讓民眾能在沒有阻礙的情況下即早發現並治療。



### **第三節 研究目的**

從動機中的兩點我們可以看到骨質疏鬆最大的困難便是難以察覺，且經常讓人無法留意，即使是經驗豐富的骨科醫師，也難以單靠個人經驗斷定確診。但骨質疏鬆卻能夠間接反應出我們的身體狀況。骨質疏鬆雖然不會直接造成我們死亡，但是卻會嚴重影響生活品質與生命安全，進而造成更多的社會問題。如果能儘早發現，不僅能降低骨質疏鬆帶來的風險，也能進一步確認自身身體狀況，多加留意會造成骨質疏鬆併發症的相關疾病。我們的研究目的便是期望用深度學習的方式來提早發覺骨質疏鬆的患者，並改善因高昂檢測費用帶來的檢測意願下降的問題，提出一種新穎且可靠又快速的解決方案。

### **第四節 研究範圍**

本研究期望利用深度學習的技術，來改進骨質疏鬆的檢測和預測方法，其中跨足包含醫學影像處理、深度學習應用、骨質疏鬆的相關診斷與臨床醫學知識，我們期望結合這些領域的發展，提出更為精確的骨質疏鬆檢測方案。

### **第五節 研究方法概述**

為了實現骨質疏鬆的預測，我們和台中榮民總醫院進行合作，並嚴格挑選樣本對象，研究對象必須鎖定有測量過 DEXA 的人群，並且還必須在前後六個月內有照過相關部位的 X 光。這些數據涵蓋了多種不同的年齡層與群體，為其確保一致性，我們也將資料進行進一步的清理與標準化，之後再以不同的深度學習模型進行嘗試，最後我們選擇 ResNet18 作為本次研究模型，接著再建立一個簡單操作的使用者介面，以提供良好的便利性。

## 第貳章 相關文獻回顧：

### 第一節 研究相關論文

在深度學習與骨質疏鬆的研究方面，我們查閱了相關的論文並將他們一一比較其差異與特點，其中在 Hsieh 等人撰寫的“Automated bone mineral density prediction and fracture risk assessment using plain radiographs via deep learning[4]”與 Chan-Shien Ho 等人發布的“Application of deep learning neural network in predicting bone mineral density from plain X-ray radiography[5]”這兩篇論文中，研究主要以預測 BMD 值為主要目標，且文中皆以大量數據和新聞證明骨質疏鬆的嚴重性，來顯示骨質疏鬆檢測模型的重要性與急迫性，第二篇文章中還提及另一個重要指標 TBS 並加以證明兩者共同判斷骨質疏鬆的效果會比以往來的更好。

而與這兩篇差異比較大的是 Yamamoto 等人撰寫的“Effect of Patient Clinical Variables in Osteoporosis Classification Using Hip X-rays in Deep Learning Analysis[6]”，這項研究旨在通過結合髋部 X 光片和患者變量(年齡、性別等)，統計評估深度學習在骨質疏鬆識別方面的能力，不同於上兩篇的地方在於，此篇直接以圖像進行骨質疏鬆的預測而非預測 BMD 值。並利用 ResNet (18, 34, 50, 101, 152) 的 CNN 模型來實作，再以 Accuracy, precision, recall, specificity, F1 score, 和 AUROC 等標準來計算模型的效能。

最後通過統計方法比較僅使用影像的模型與集成模型（影像和患者因素）的準確度，並進行統整。

### 第二節 架構相關論文

在架構方面我們根據以往相關的研究，選擇以 ResNet 作為實驗模型，並比較不同種類的模型在骨質疏鬆判別上的成果，如 GoogleNet、DenseNet、VGG16 等不同模型在表現上與 ResNet 的差別。而因為 GoogleNet 在實驗中和 ResNet 有類似的表現，所以我們特別介紹一下 GoogleNet 和 ResNet 的模型技術。

ResNet 的核心理論在這篇由 Kaiming He 等人發表的“Deep Residual Learning for Image Recognition[7]”中指出，testing error 和 training error 同樣因為層數的加深反而變得更高的狀況稱之為“degradation”。這種有別於 Overfitting 的狀況便是因為卷積層

(convolution layer)的持續增加所導致的。為了解決這個問題，文中提出以學習殘差塊的方式來減緩梯度消失的想法，殘差塊的核心思想是引入了跳躍連接（skip connection），允許淺層的輸出直接添加到深層的輸入上。這樣能讓電腦更輕鬆的學習淺層的低級特徵，從而減緩因深度增加而造成的特徵流失[8]。

而 GoogleNet 又稱為 Inception 架構，是由 Christian Szegedy 等人提出的方法，並撰寫這篇名為“Going deeper with convolutions[9]”的論文。這個架構的主要特點是使用了一個稱為 Inception 模塊的結構，讓電腦能夠有效的提取和組合圖像的不同特徵。在 Inception 模塊中，將訓練過程分成多個分支，並分別使用了多種不同尺寸的卷積核，這種多分支結構讓電腦能夠在不同大小的卷積上進行特徵提取，並在不同層中組合這些特徵，從而提高了網絡對判別不同特徵的能力[10]。

### 第三節 方法相關論文

現行骨質疏鬆檢測中，國際骨質疏鬆協會認可了以 DEXA 骨密儀測量的 BMD 作為重要骨質疏鬆的參考依據，另外還有像 TBS 這種較為新型的骨質疏鬆指標，在這篇由 N.C. Harvey 等人撰寫的論文“Trabecular bone score (TBS) as a new complementary approach for osteoporosis evaluation in clinical practice[11]”中，證明了 TBS 如何影響骨質疏鬆，這種新型指標 TBS 是通過重新分析 DEXA 影像來獲得，從量化 DEXA 所生成的圖像數據，探討從一個像素到相鄰像素的灰度級紋理的變化。

而在 2008 年的國際骨質疏鬆會議上，更是確定了 DEXA 骨密儀的重要性與地位，在這篇“A reference standard for the description of osteoporosis[1]”的會議記錄文中，提出了幾點對我們此次研究非常重要的依據：

1. 提出了骨質疏鬆的黃金標準---“骨礦物質密度(BMD)”，並採用雙能量 X 射線(DEXA)進行測量。
2. 並以 20~29 歲骨質發育最完整的女性股骨頸平均 BMD，做為 BMD 的標準值。
3. 進一步說明骨質疏鬆的定義為低於此標準值 2.5 個標準差的患者。
4. 並指出任何新的檢測方式都應該以雙能量 X 射線(DEXA)作為參考標準，以確保新技術足夠可靠且穩定。

我們的實驗也會以骨質疏鬆協會的規定，以雙能量 X 射線(DEXA)做為參考標準，來證明我們的實驗成果是否有效。

## 第參章 研究流程擬定：

### 第一節 研究方法

本論文所開發的「骨質疏鬆檢測系統」，目標是希望透過深度學習的方式以及利用 X 光影像高解析度之優勢來進行骨質疏鬆的診斷，因此我們將會利用卷積神經網絡 (Convolutional Neural Network) 即 CNN 的方法去探討上述目標之可能性，在模型架構上，我們會以 ResNet18 作為我們的訓練模型並且對髖關節進行圖像分類 (image classification)。

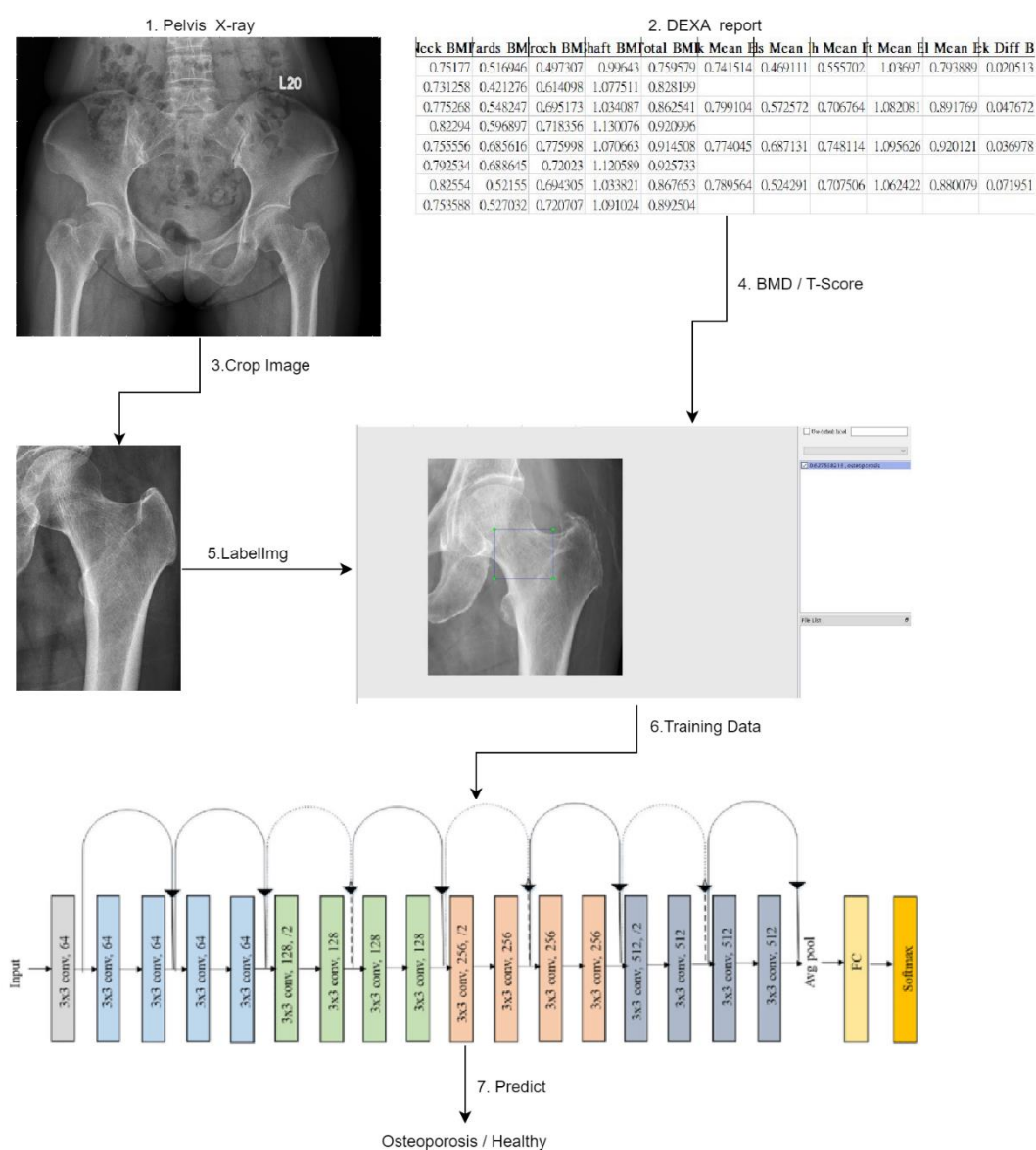


圖 3-1 流程圖[12]

## 第二節 方法架構

本系統的整體架構圖如上圖 3.1 所示，此外會結合由 Django 所建立的網頁去進行骨鬆檢測，讓使用者可以直接從網站上上傳自己股骨的圖片，再經由已訓練好的深度學習模型進行骨鬆的診斷，最後再把診斷結果回傳到使用者的介面。

如本系統診斷出使用者患有骨質疏鬆症，主程式將會向使用者發出警告信息，提醒使用者要去醫院作進一步檢查之必要性。同時主程式也會結合 Grad-CAM 所產生的熱度圖去顯示系統所偵測到骨鬆的位置。

在研發流程上，「骨質疏鬆檢測系統」的研發流程主要分成四大部分，分別包括資料收集與前處理，圖像增強，訓練用於診斷骨質疏鬆的深度學習模型，以及完成訓練後結合模型之 API。步驟如下：

### 1. 資料收集與前處理：

首先，在資料收集與前處理的部分，我們從台中榮總醫院收集到從 2018 年到 2022 年間有拍過髌部 X 光和測量雙能 X 射線吸收骨密度儀 (DEXA) 的病人的資料。而在這些資料當中，會有一些病人只拍過髌部 X 光卻沒有測量過 DEXA，因此我們需要利用資料庫去進行資料的篩選和病人的配對。

對於資料的篩選，為了確保訓練資料的正確性和代表性，圖片篩選的條件非常嚴謹，病人需要同時拍過髌部 X 光以及測量過 DEXA，而且兩者之間的間隔不能相差半年或以上，換句話說，就是病人在測量 DEXA 過後的前 6 個月內或者後 6 個月內所拍的髌部 X 光片才能成為我們的訓練對象。為了能夠大量以及快速的篩選資料，我們使用 Oracle 資料庫來進行資料的配對與篩選。

首先，我們從醫院所收集到的 DEXA 測量數據是 CSV 格式的檔案而髌部 X 光圖片則為 Dicom 的檔案格式。簡單介紹，Dicom 檔案是目前世界上所制定的醫療圖像儲存、交換及傳輸的一套標準。除了儲存影像外，還會記錄病人的基本和檢查等資料，例如病人的姓名、Patient ID、身高、體重等等的資訊，由於 Dicom 檔案無法直接從電腦上開啟，因此我們首先需要利用 python 函式庫中的 Pandas 和 Pydicom 來把這些數據讀取並轉換成 DataFrame，並輸出成 CSV 檔。得到了髌部 X 光的 CSV 檔後，就可以將這些 CSV 檔導入到 Oracle 的資料庫中並且利用 Patient ID 的唯一性把前後 6 個月內拍過髌部 X 光和測量過 DEXA 的病人配對出來。成功配對的數據後，最後就會把這些髌部 X 光圖片利用正規化的方式全部轉換出來，並儲存成 PNG 格式。

接下來，就是標註的部分，我們會使用 LabelImg 對髖部中股骨頸的部位進行標註，如下方的圖 3.2 所示，標註的依據是透過仿照 DEXA 機器在測量髖部 BMD 值時所量測之部位，同時也經過了台中榮總醫生審視和同意而得出的結果，通過對照 DEXA 數據中的 Neck\_BMD 值和 T-score 來把資料分成兩個類別，分別是有骨質疏鬆症(osteoporosis)和沒有骨質疏鬆症(osteopenia/healthy)，此外在標註的過程中，我們也會把不符合標準的圖片進行再一輪的篩選，例如清晰度欠佳的圖片，或是在股骨上有金屬支架等物體，見圖 3.3，這些外在因素有可能會導致訓練時的誤差，因此我們先會排除這類型的圖片，避免影響訓練效果。



圖 3-2 標註部位



圖 3-3 金屬支架



圖 3-4 歪斜情況

## 2. 圖像增強：

經過資料的配對、標註與篩選後，我們總共得到了 396 張圖片。我們會把 2018 到 2021 年間的圖片作為訓練集以及驗證集的分割，共有 312 張，而將 2022 年的資料作為測試集以模擬真實上線的情況，共有 84 張。進一步我們利用圖像增強的方式去進行資料的擴增。具體來說，我們將訓練集的圖片進行隨機旋轉( $-20^{\circ}$ ~ $20^{\circ}$ )，然後會有 30% 的機率會把圖片進行隨機光暗度調整( $-5^{\circ}$ ~ $5^{\circ}$ )。其中隨機旋轉的目的是為了模仿一些病人因有骨頭上的其他疾病而導致歪斜的情況，如上方的圖 3.4 所示，同時光暗度調整是為了防止模型在訓練時過度依賴光暗度作為判斷因素而作出的一些干擾。

同時，我們也將右邊的股骨進行水平反轉，使其成為左股骨資料集的一部分，同樣的我們也對右股骨進行了配對、標註與篩選的操作，共得到 297 張的有效圖片並且全部會用作訓練集中一部分，此外，我們對右側股骨也進行了同樣的圖像增強方式。經過圖像增強後訓練集從 312 張擴增至 1218 張。

### 3.訓練用於診斷骨質疏鬆的深度學習模型：

經過了前處理以及圖像增強後，接下來進入訓練模型的部分。本次研究會以交叉驗證(K-fold Cross-Validation)，將訓練集隨機分配到五個集合之中進行交叉驗證，每次訓練時 epoch 設定為 50，batch size 設定為 5，Optimizer 選擇 SGD，learning rate 設定為 0.001，同時選擇了 BCEWithLogitsLoss 作為損失函數。此外，我們將骨質疏鬆症和沒有骨質疏鬆症的標籤轉換成 0 跟 1，其中 1 代表有骨質疏鬆症，而 0 就是代表沒有骨質疏鬆，其中每一張圖片都會 resize 成 224\*224。

最後利用 ResNet18 模型架構進行訓練。完成訓練後，我們選擇了效果最好的一次作為我們的系統的檢測系統，並把該次訓練的結果備份成 pth 檔。

### 4.訓練後結合模型之 API：

為了提供更好的便利性，我們會透過網站的形式來實現我們系統的 API，我們將使用 Django 來結合深度學習模型來實現骨質疏鬆檢測系統的 API。Django 是一個基於 Python 程式語言的 Web 開發框架[11]，同時提供了幾乎任何格式的內容，例如 HTML、CSS 等。而本系統的操作流程如圖 3.5 的 Django 架構圖所示。當用戶在瀏覽器上傳股骨的圖片時，它會向 Django 伺服器發送一個 HTTP 請求。而當 HTTP 請求到達伺服器時，urls.py 就會根據請求的 URL 路徑來決定哪一個視圖(View)來處理這個請求。這個時候就會透過相對應的 View 與訓練好的骨質疏鬆檢測模型及 Grad-Cam 程式進行結合。最後生成的 HTML 頁面會作為 HTTP 回應返回給使用者的瀏覽器，使得使用者在網站上傳股骨的圖片後，就能得到這兩個主程式回傳的結果。進而達到使用者介面之功能。

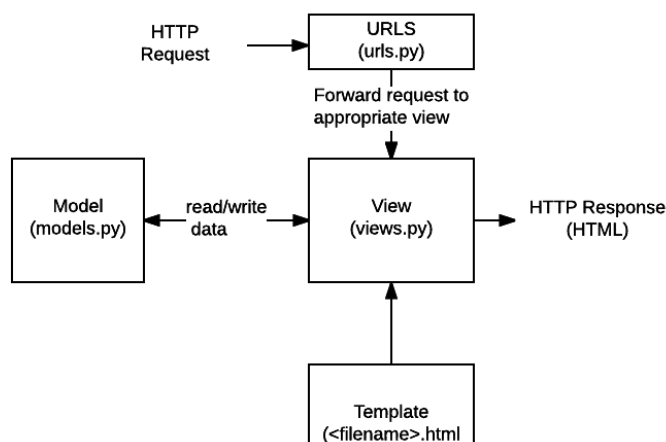


圖 3-5 Django 架構圖[13]

## 第肆章 實驗過程論述：

### 第一節 實驗流程

論文將會以四階段的實驗來論證第三章所提出之方法的正確性、適當性與可行性，並同時提出數據進行比較。

#### 第一階段實驗

在第一階段實驗，我們會先以原始數據來進行模型的訓練，不加上資料增強的步驟，目的是為了觀察模型對於原始數據訓練時之效果，因此訓練資料為2018到2021年的股骨 X 光圖片，共 312 張的圖片。而模型架構同樣會先以 ResNet18 進行訓練，epoch=50，batch size=5，Optimizer 選擇 SGD，learning rate 設定為 0.001，同時選擇了 BCEWithLogitsLoss 作為損失函數。而在資料切分方面，會使用 random\_split 的方法把資料隨機分成 80% 和 20% 分別是訓練集以及測試集。從結果中可以發現，只利用原始數據來訓練的效果並不理想，在圖 4.1 的效能圖中可以看到，除了準確率非常低以外只有 73% 的準確率，同時在圖 4.2 中 Loss 的折線圖也可以發現，模型出現了 overfitting 的問題，而我們認為是訓練資料量不足所導致模型無法收斂。

Classification Report:				
	precision	recall	f1-score	support
0	0.71	0.85	0.77	34
1	0.77	0.59	0.67	29
accuracy			0.73	63
macro avg	0.74	0.72	0.72	63
weighted avg	0.74	0.73	0.72	63

圖 4-1 效能圖

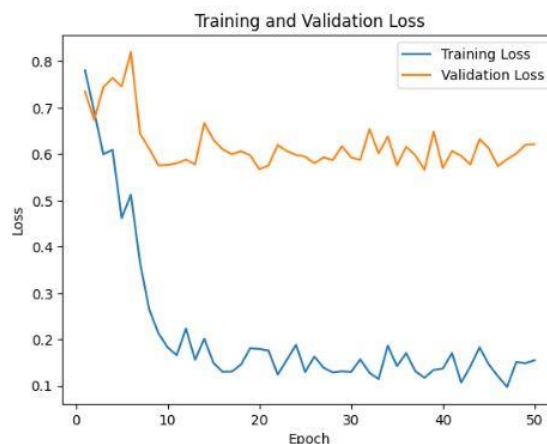


圖 4-2 折線圖



## 第二階段實驗

經過了第一階段的實驗過後，我們發現資料的缺乏導致 overfitting 的問題，因此我們會利用圖像增強的方法來進行訓練集的擴增，利用第三章所提出的方法進行圖像增強，將資料從 312 張擴增至 1218 張。同樣會使用跟第一階段一樣的模型架構、學習率和損失函數進行訓練。在圖 4.3 和圖 4.4 中，從結果上可以看到，overfitting 的問題得到了大幅的改善，而準確率也有明顯的提升。

Classification Report:				
	precision	recall	f1-score	support
0	0.89	0.87	0.88	135
1	0.84	0.87	0.86	109
accuracy			0.87	244
macro avg	0.87	0.87	0.87	244
weighted avg	0.87	0.87	0.87	244

圖 4-3 效能圖

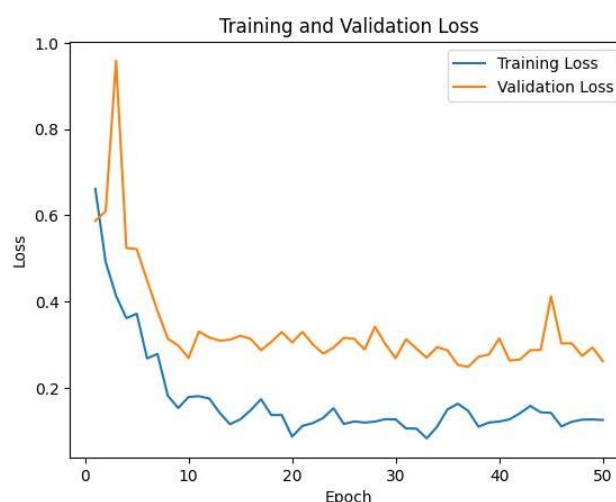


圖 4-4 折線圖

## 第三階段實驗

接下來，我們統整過去相關論文中出現的各種模型將會通過測試不同的模型架構來探討哪一種模型的效果是最好的，我們會嘗試以 ResNet18, ResNet34, GoogLeNet, DenseNet 和 VGG16 來做深入比較。其中我們會以執行時間、驗證集 Loss、Accuracy 和 Recall 等作為我們的衡量標準。為了能更全面的評估模型的能力，本論文將會使用交叉驗證(K-fold Cross-Validation)來進行資料的切分，K 會設定為 5，因此每個模型會個別進行 5 次的訓練，並在結束訓練後把 5 次的驗證集 loss 加總取平均，以此來評估不同模型之間的效果的 val\_loss 平均值(圖 4.5)。

為了更進一步的測試模型的泛化能力，我們會將每個模型在五次的訓練當中把訓練效果最好的一次記錄下來，並儲存成 .PTH 檔，接著，我們會以 2022 年的資料作為測試集以模擬真實上線時的情況進行預測，並會以 Classification report (分類報告) 和 Confusion matrix (混淆矩陣) 觀察模型性能。

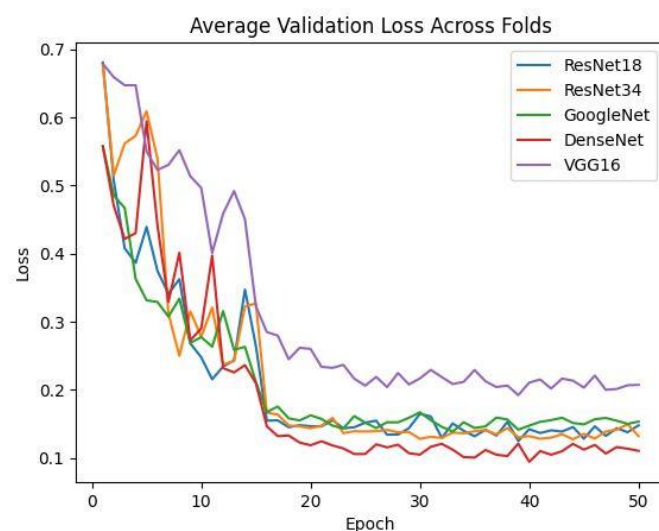


圖 4-5 Loss 圖表

表格 4.1 各模型對於無骨質疏鬆的測試集數據

模型種類	Precision	recall	f1-score	time(min)
ResNet18	0.81	0.81	0.81	30.6
ResNet34	0.83	0.78	0.81	55
GoogLeNet	0.79	0.84	0.82	52.4
DenestNet	0.73	0.89	0.80	114.3
VGG16	0.68	0.92	0.78	157.7

表格 4.2 各模型對於有骨質疏鬆的測試集數據

模型種類	Precision	recall	f1-score	time(min)
ResNet18	0.85	0.85	0.85	30.6
ResNet34	0.84	0.87	0.85	55
GoogLeNet	0.87	0.83	0.85	52.4
DenestNet	0.90	0.74	0.81	114.3
VGG16	0.91	0.66	0.77	157.7

首先，在驗證集的圖 4.5 Loss 圖中可以發現，除了 VGG16 以外，每個模型在驗證集中都有非常不錯的表現，而且在預測驗證集的準確率也能平均達到接近 9 成，說明了利用股骨來進行影像分類是可行的做法。然而在 2022 年的測試集表格 4.1 和表格 4.2 中發現，當模型面對真實的資料時，準確率上的表現卻會明顯下降，這可能是因為資料量的缺乏導致的問題，最後使模型的泛化能力下降。

其次，在預測 2022 測試集中的 Classification report 發現，表現最好的模型分別有 ResNet18, ResNet34 和 GoogleNet，其準確率（Accuracy）都一樣為 0.83。然而，除了 Accuracy 以外，Recall 也是本論文衡量模型的重要指標之一。由於此次研究的關鍵是希望把潛在的骨質疏鬆患者找出來，因此當 Recall 很低的時候就會有較高的機率將骨質疏鬆的病人誤判成沒有骨質疏鬆，這就與我們研究的方向背道而馳，所以這些 Recall 指標成了最重要的比較對象。從上述列表可以看出 DenseNet 和 VGG16 在 Recall 上的表現並不理想，因此在後續的訓練中並不會將它們納入考慮。

經過多次重覆的訓練，發現導致兩個類別的 Recall 差異很大，我們認為是樣本偏差所導致，跟據我們的統計，在 312 筆原始左股骨數據中，有骨質疏鬆的樣本有 141 筆，而沒有骨質疏鬆的樣本則為 171 筆，相差 30 筆，同時在 297 筆原始右股骨數據中，有骨質疏鬆的樣本有 124 筆，而沒有骨質疏鬆的樣本則為 173 筆，相差 49 筆，經過資料強化後兩類別相差了 158 筆，因此每次訓練時的隨機切分都有可能造成類別上分配不均，所以我們並不能單靠一張的性能表直接斷言出哪一個模型對於 Recall 有較好的表現，但是能夠看出相較於 DenseNet 和 VGG16，ResNet 和 GoogLeNet 是對於骨鬆影像分類的任務有更高的潛力。

總而言之，透過實驗三我們探討了各模型的效能以及點出資料分配不均的潛在問題，對於後續的模型調整和實驗，我們會選擇使用 ResNet18，其原因是 ResNet18 有較高的 Recall 之外同時也有較快的執行時間，這樣可以方便我們在後續實驗中調整不同的參數進行嘗試。

## 第四階段實驗

為了解決類別不平均的問題，我們嘗試了欠採樣 (Undersampling)，但我們發現在資料已經缺乏的情況下，隨機刪除部分樣本並不是明智的做法，因此在嘗試過後雖然可以解決到類別不平均的問題，但實際上準確性並沒有得到明顯的提升，所以我們認為要從根本上解決問題就只能增加訓練的資料量。

除此之外，我們也有對訓練的各個超參數進行調整，我們將學習率調度器 StepLR 中的 step\_size 從原來的 7 改成 15，讓模型在訓練 15 個 epoch 後會更新一次學習率，我們發現這對於模型的收斂效果會變得更明顯，從結果來看準確率也得到提升(圖 4.6 和圖 4.7)，而這些都是經過多次實驗調整所發現的結果，可以通過對比實驗二來發現之間的差異。

Classification Report:				
	precision	recall	f1-score	support
0	0.97	0.93	0.95	131
1	0.92	0.96	0.94	112
accuracy			0.95	243
macro avg	0.95	0.95	0.95	243
weighted avg	0.95	0.95	0.95	243

圖 4-6 效能圖



圖 4-7 折線圖

## 第二節 實驗結果

經過了四階段的實驗後，我們探討了我們在訓練時所遇到的各種問題，以及提出了各種的解決方案。最後我們將會利用訓練效果最好的一次作為我們系統的檢測模型。圖 4.8 這個結果是基於預測 2022 測試集的效能圖，可以看到準確率達到了 0.92，而骨鬆類別的 recall 達到了 0.96，我們對於這個結果是滿意的。

Classification Report:				
	precision	recall	f1-score	support
0	0.94	0.86	0.90	37
1	0.90	0.96	0.93	47
accuracy			0.92	84
macro avg	0.92	0.91	0.91	84
weighted avg	0.92	0.92	0.92	84

圖 4-8 結果效能圖

## **第伍章 結論：**

### **第一節 結論**

骨質疏鬆不同於一般的疾病，無法單靠醫師的經驗來判斷病患的嚴重程度，而是需要透過精密醫療設備的幫助，然後再以國際骨質疏鬆協會的定義來間接判斷病人的狀況，這樣不只延長了治療上的時間，也需要花費更多的成本和人力，同時也降低了病患就診的意願。

總結而言，我們希望提供的是一種新型且現代化的解決方案，不僅能提供更低的价格和更短的時間，同時也提供與現今方法相同可靠的服務，我們期望的不只侷限在價格上的不同，我們更期待它能成為新一代的主要骨質疏鬆治療方案，讓被稱為沉默殺手的骨質疏鬆疾病問題從根本被解決。

### **第二節 研究限制討論**

#### **1.數據限制**

本次研究在實驗上最大的限制便是數據取得不易，醫療相關資源與照片存在倫理與個人隱私等問題。加上為求實驗精確且有效，在嚴格挑選樣本的同時，過濾掉九成以上的影像檔案，如能有更完整且充足的數據進行訓練，一定能得到更好的效果。

#### **2.時間限制：**

從 4 月初的題目訂定到現在經過約 6 個多月的研究，看似充足，實則非常緊迫，從骨質疏鬆的相關定義論文到相似研究的探討、部位的選擇、訓練遇到的瓶頸和最後的論文撰寫等等。而我們能壓縮的部分便是不同部位上的選擇，我們只能透過相關的研究，選出可行性最高的股骨頸做為實驗對象，如能有充裕的時間，我們也想嘗試不同部位，如腰椎或胸腔等等的 X 光片對於判別骨質疏鬆是否也同樣可行，如果增加不同部位對於骨質疏鬆的研究，便能進一步使骨質疏鬆的判別變得更加容易和方便。

### 3.樣本選擇偏差：

數據的取得來源為台中榮民總醫院，因為測量 DEXA 需要自付一定費用且須預約候診，在這樣的原因下很大程度的排除了健康的 DEXA 資料，由於幾乎都是有一定的症狀或是有相當大的疑慮才會進行 DEXA 的測量，這樣導致了分類上的不容易與不同類別之間存在基數差距過大的問題。如資料分部能更加平均，我們希望能將預測結果分成三類，讓骨質疏鬆初期的患者也能被檢測出來，並在第一時間得到妥善的預防性治療。

## 第三節 研究成果

此次研究的成果在預測 2022 年的測試集中準確率和召回率都能夠達到 9 成以上，這證明了股骨 X 光影像的圖像分類對於骨鬆領域是一個可行的方法，再配合使用者介面讓使用者方便操作。圖 5.1 是本系統的主頁，使用者可以在網站上直接上傳股骨圖片進行預測，此外也附上了使用者手冊教導使用者如何使用本系統，當使用者點擊使用者手冊的時候，畫面就會跳轉至使用者手冊的介面圖 5.2。

圖 5.3 是當本系統的預測到骨鬆時的畫面，網頁會顯示警告信息，同時結合 Grad-Cam 來顯示系統偵測到骨鬆的位置，此外使用者也可以透過點擊進行預約來進行網絡掛號，本論文以台中榮民總醫院為例子，圖 5.4。最後附上骨質疏鬆自我評量表供使用者參考，基於本論文採用了創新的方法來測量骨質疏鬆，了解到可能會引來各方的質疑，因此希望透過評量表的方式來提升本系統的可信度，同時也希望鼓勵落在高風險區的使用者可以去醫院作進一步的檢查，達到了推廣的作用。

## 第四節 研究貢獻

我們統整了過去相關的論文研究，不論是在模型上的選擇或是預測方式，各種不同路線的研究方式我們都有嘗試過，在互相比較不同方法的同時提出我們自己的看法，並在對模型的研究成果上，建立一個簡單的使用者介面，能讓即使沒有相關知識背景的人，也能更加直覺的感受到我們在骨質疏鬆領域上的研究成果。

## 第五節 未來展望與後續研究

此次研究的系統距離可以真實上線其實還有一段距離，其原因是能夠訓練的資料量相對其他研究較少，我們希望未來可以與更多的醫院合作取得更多的 X 光與 DEXA 數據來完善我們的系統，目前在大概 500 張資料的情況下也有如此不錯的效果，其實已經證明了它的潛力，我們非常希望本系統未來能正式上線並成為一個可靠和主流的骨鬆檢測方法。

以下是實際操作畫面：



圖 5-1 網站主頁



圖 5-2 使用者手冊介面

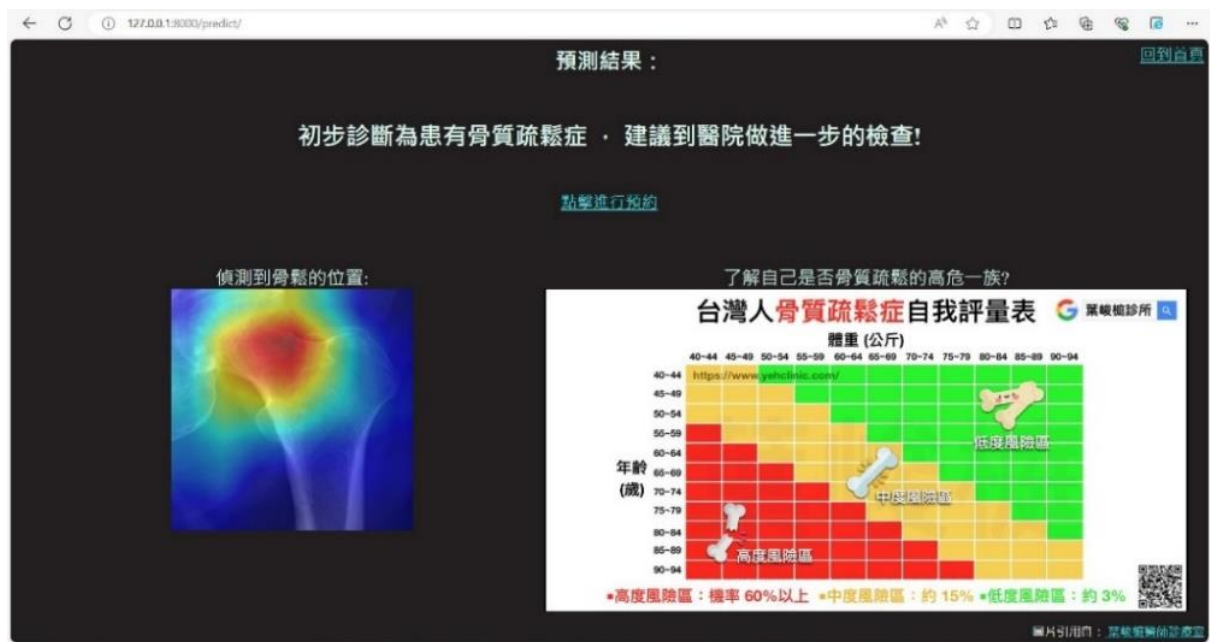


圖 5-3 偵測到骨質疏鬆並附上自我評量表[14]



圖 5-4 網絡掛號



## 參考文獻

- [1] John A. Kanis, Eugene V. McCloskey, Helena Johansson, Anders Oden, L. Joseph Melton, Nikolai Khaltsev, "A reference standard for the description of osteoporosis," *Elsevier*, Mar 2008
- [2] DEXA 價格: <https://dexaplus.com/dexa-scanner-cost/>
- [3] 衛生福利部相關報導: <https://www.mohw.gov.tw/cp-2632-14705-1.html>
- [4] Hsieh, C.I., Zheng, K., Lin, C. et al. "Automated bone mineral density prediction and fracture risk assessment using plain radiographs via deep learning," *Nat Commun*, Sep 2021.
- [5] Ho, C.S., Chen, Y.P., Fan, T.Y. et al. "Application of deep learning neural network in predicting bone mineral density from plain X-ray radiography," *Arch Osteoporos*, Oct 2021.
- [6] Yamamoto, N.; Sukegawa, S.; Yamashita, K.; Manabe, M.; Nakano, K.; Takabatake, K.; Kawai, H.; Ozaki, T.; Kawasaki, K.; Nagatsuka, H. et al. "Effect of Patient Clinical Variables in Osteoporosis Classification Using Hip X-rays in Deep Learning Analysis," *Medicina*, Aug 2021.
- [7] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Sep 2016.
- [8] 論文閱讀:ResNet(Deep Residual Learning for Image Recognition):  
<https://hackmd.io/@RinHizakura/HJ7rNmTV8>
- [9] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich, "Going Deeper with Convolutions," N/A, 2014.
- [10] Review: GoogLeNet: <https://medium.com/@chensheep1005/googlenet-ac609d13e3f1>
- [11] N.C. Harvey, C.C. Glüer, N. Binkley, E.V. McCloskey, M.-L. Brandi, C. Cooper, D. Kendler, O. Lamy, A. Laslop, B.M. Camargos, J.-Y. Reginster, R. Rizzoli, J.A. Kanis, "Trabecular bone score (TBS) as a new complementary approach for osteoporosis evaluation in clinical practice," *Elsevier*, Sep 2015.
- [12] ResNet18 架構圖:  
[https://www.researchgate.net/figure/Original-ResNet-18-Architecture\\_fig1\\_336642248](https://www.researchgate.net/figure/Original-ResNet-18-Architecture_fig1_336642248)
- [13] Django 介紹以及架構圖:  
<https://developer.mozilla.org/zh-TW/docs/Learn/Server-side/Django/Introduction>
- [14] 骨質疏鬆自我評量表: <https://www.yehclinic.com/osteoporosis-screening/>