**結合大語言模型之醫療影像辨識模型與系統開發—以骨科為例**

摘要

本研究針對一般民眾不易理解醫療影像的問題，以骨科 X 光片為切入點，建置「骨骼健康教育系統」，作為提升骨骼知識與輔助教學的工具。系統包含四大核心模組：影像標記、影像辨識、知識檢索與聊天互動。技術上整合影像辨識模型、大語言模型、檢索增強生成技術與前後端網頁技術，目標是產生具有高可讀性、可信度與易於理解的骨骼知識內容。目前初步結果顯示，影像辨識正確率達到91.6%，證實模型對骨骼位置與類別的判斷具有高度準確性與可靠性，是系統進行影像辨識及骨骼健康教育的關鍵基礎。

關鍵詞：醫療影像辨識、骨科、YOLO、大語言模型、健康教育

# 

# 第 1 章　緒論

1.1 發展背景與動機

深度學習在醫療影像分析領域已被廣泛應用，從腫瘤偵測到器官分割都有相當成熟的研究成果，相關綜論指出，卷積神經網路在多種影像模態上皆展現優異表現（Litjens 等，2017）。在骨科領域，深度學習也逐漸用於骨折偵測與骨骼結構分析，協助醫師加速判讀 X 光影像並提升診斷一致性（Su 等，2023；Lee 與 Chung，2022）。然而，多數研究主要著重於輔助醫師或放射科人員判讀，對於如何幫助病人本身理解自己的影像，仍較少有完整系統設計。

在實際看診情境中，一般民眾面對 X 光影像時，通常只能看到黑白陰影與醫師快速的口頭說明，很難形成「這一塊是什麼骨頭」「這個位置平常負責什麼動作」「受傷大概會有哪些影響」等具體理解。市面上與學界已有不少三維骨骼教學系統，透過 3D 模型、旋轉與拆解等互動方式，協助使用者理解骨骼結構，部分系統是由教育部或其他機構支持的解剖學教學平台，本文以這類系統統稱為「教育部 3D 骨骼教學系統」。這些系統在標準化模型教學上具有優勢，但多半以標準人體模型為主，與個人 X 光影像之間缺乏直接連結，也較少結合深度學習偵測與大語言模型進行即時文字解說。

近年大型語言模型（Large Language Model, LLM）在自然語言理解與生成上的能力已有顯著突破，GPT-4 技術報告顯示，模型在多項測驗上達到接近人類的水準（Achiam 等，2023）。同時，檢索增強生成架構（Retrieval-Augmented Generation, RAG）提出，在生成答案前先從外部文件中檢索相關內容，可減少模型憑空想像的問題（Lewis 等，2020）。如果能將骨科影像偵測與 LLM＋RAG 結合於同一系統中，便有機會在「病人的實際影像」與「可理解的文字解說」之間建立橋樑。

為方便後續說明，本研究將整體系統統一命名為「骨骼健康教育系統」（Orthopedic Bone Health Education System, OBHES），並依功能區分為三個模組。OBHES 影像偵測模組（Image Detection Module, OBHES-IDM）負責接收上傳之 X 光影像，執行骨骼區域偵測並產生對應骨頭編號與信心度；OBHES 知識檢索模組（Knowledge Retrieval Module, OBHES-KRM）整合骨骼結構資料與教材內容，採用 RAG 流程擷取與問題相關的文字片段；OBHES 聊天模組（Intelligent Explanation and Dialogue Module, OBHES-IEDM）則結合檢索結果與使用者問題，利用大語言模型產生一般民眾可理解之自然語言解說與問答。

### 1.2 研究目的

本研究以建置「骨骼健康教育系統」為目標，針對一般民眾與教學場域在理解骨科 X 光影像上的困難，提出一套結合影像標記、影像辨識、知識檢索與聊天互動四個核心模組的整合性解決方案。首先，在資料端的角度，本研究希望透過影像標記模組建立一套可重複使用的標註與回饋機制，讓專業人員能對骨科 X 光影像進行骨骼位置與類別的人工標註與修正，形成後續模型訓練與效能評估所需的高品質資料基礎。第二，在影像處理層面，本研究致力於設計影像辨識模組，讓使用者能透過網頁介面上傳骨科 X 光影像，由物件偵測模型自動偵測骨骼區域並輸出骨頭編號與可信度，並與後端骨骼結構資料表連結。第三，在知識層面，本研究規劃知識檢索模組，將骨骼結構化資料與教材內容整合為可檢索的知識來源，未來可搭配檢索增強生成流程，根據特定骨頭與問題擷取相關文字片段。第四，在使用者互動層面，本研究設計聊天互動模組，將使用者問題、偵測到的骨頭編號與知識檢索結果組合為提示語，交由大型語言模型產生一般民眾與學生可理解的自然語言解說與問答，並透過對話紀錄累積系統實際使用情形，以作為後續評估與優化之依據。

### 1.3 研究範圍

本研究的範圍以骨科 X 光片為主要資料來源，聚焦於胸廓區域之骨骼名稱與位置說明，並據此設計影像標記、影像辨識、知識檢索與聊天互動四個核心模組及相關資料結構。整體架構以 X 光影像為起點，在設計上亦預留擴充空間，未來可擴充延伸至其他醫療影像型態與更多骨骼部位。

# 第 2 章　文獻探討

### 2.1 醫療影像標註的重要性

在醫療影像領域，高品質標註被視為深度學習模型的前提，諸論指出醫療影像的標註成本高且需專業人員介入，是實務導入的主要瓶頸之一（Litjens 等，2017；Su 等，2023）。骨科相關研究也強調，骨折與骨骼偵測效能高度依賴一致且精確的標註資料，若標註含糊或標準不一，模型在臨床情境下的可靠度將明顯下降（Alzubaidi 等，2024）。因此，本研究設計影像標記模組，不僅用於產生訓練資料，也提供持續修正與補充標註的機制，以支撐系統後續的模型更新與教學內容擴充。

### 2.2 醫療影像與骨科領域之深度學習應用

深度學習導入醫療影像分析以來，已在腫瘤偵測、肺部結節分析與視網膜病變篩檢等領域產生顯著影響。綜論指出，卷積神經網路在影像分類與分割任務上表現優異，並逐漸被考慮納入臨床輔助決策流程（Litjens 等，2017）。在骨科領域，Su 等（2023）整理骨折偵測研究，指出多數方法以 X 光影像為主要資料來源，將骨折問題表述為分類、偵測或區域定位等不同型態，並強調資料集規模與標註品質的重要性。Lee 與 Chung（2022）則從骨科疾病整體角度出發，回顧深度學習在退化性關節疾病、脊椎異常與人工關節分析中的應用，顯示骨科領域對人工智慧的需求持續增加。

### 2.3 物件偵測模型與 YOLO 在骨折偵測上的應用

物件偵測模型同時預測影像中物件的位置與類別，常見方法包括兩階段偵測器與單階段偵測器。You Only Look Once 物件偵測模型（You Only Look Once, YOLO）屬於單階段方法，將偵測視為單一回歸問題，在一次前向運算中完成邊界框與類別機率的預測，兼具速度與準確度。Ju 與 Cai（2023）將 YOLOv8 應用於兒童手腕創傷 X 光影像的骨折偵測，訓練結果顯示，在適當的資料前處理與模型設定下，YOLO 模型能達到具有臨床輔助價值的偵測效能。Srinivasu 等（2025）則以 YOLOv10 探討超參數與資料增強對骨折偵測準確度與平均準確率的影響，指出在骨科影像任務中，模型架構與訓練細節皆會左右最終成果。這些研究顯示，YOLO 架構在骨折與骨骼偵測上具有良好基礎，適合作為本研究 OBHES-IDM 的核心演算法。

### 2.4 大語言模型與檢索增強生成

大型語言模型在語言理解與生成上的能力，為專業領域的問答與解說系統帶來新的可能性。GPT-4 技術報告指出，模型在多項標準化測驗與專業考試中表現亮眼，但也必須注意模型可能產生與事實不符的幻覺回答（Achiam 等，2023）。為降低此風險，Lewis 等（2020）提出檢索增強生成架構（Retrieval-Augmented Generation, RAG），透過事先對文件向量化與索引，讓系統在生成答案前先檢索出最相關的文本片段，再交由語言模型參考後生成回答。這種方式在知識密集型任務中展現出比純生成模式更穩定的表現。本研究於 OBHES-KRM 中採用類似概念，將骨骼結構資料與教材文本轉換為可檢索的文件集合，讓 OBHES-IEDM 在生成解說文字前，先以 RAG 擷取與特定骨頭與問題最相關的內容，以提升回答的一致性與可信度。

### 2.5 現有 3D 教學產品與本研究定位

市面上已有多款三維解剖學與骨骼教學產品，包含由教育部或國際出版社支持的 3D 解剖學平台，使用者可透過高精細 3D 模型、分層顯示與互動操作建立空間概念。以 Motsinger（2020）介紹的 Complete Anatomy 為代表，其提供完整人體 3D 解剖模型與多樣互動功能，已廣泛應用於醫學教育與自學情境。相關醫學教育研究指出，這類 3D 工具對理解複雜解剖結構具有正面效果，特別是對初學者而言，立體視覺化能補足平面圖譜的不足。然而，無論是國內教育部支持之 3D 骨骼教學系統，或國際 3D 解剖學平台，多數仍以標準人體模型為主，較少處理病患個人影像，也未必結合深度學習影像偵測或大語言模型進行即時文字解說。

　　相較之下，本研究提出的 OBHES 在系統定位與技術設計上有不同的著力點。首先，OBHES 影像偵測模組（OBHES-IDM）直接處理使用者上傳之實際骨科 X 光影像，而非僅依賴標準 3D 模型，強調個人影像與骨骼名稱、位置之連結。其次，OBHES 知識檢索模組（OBHES-KRM）與聊天模組（OBHES-IEDM）結合檢索增強生成（RAG）與大型語言模型，從骨骼結構資料與教材中擷取與特定骨頭及問題相關的內容，產生一般民眾較易理解的文字解說與問答。換言之，本研究更關注「如何讓使用者讀懂自己的影像」，而非僅在標準人體模型上建立抽象的空間概念。

　　為更清楚說明本研究與現有產品及文獻中系統的差異，表 2-1 以「教育部 3D 骨骼教學系統」「國際 3D 解剖學平台（以 Complete Anatomy 為例）」與本研究 OBHES 為比較對象，從用途、資料來源、技術核心與互動方式等面向進行整理。

表1　現有 3D 系統與本研究 OBHES 之比較

|  |  |  |  |
| --- | --- | --- | --- |
| **比較面向** | **教育部 3D 骨骼教學系統** | **國際 3D 解剖學平台（Complete Anatomy,Motsinger, 2020）** | **本研究 OBHES** |
| 主要用途 | 提供國內課程用之骨骼與解剖學標準化教學 | 提供完整人體 3D 解剖學學習環境，支援課堂與自學 | 協助一般民眾與學生理解自己 X 光影像中的骨骼名稱與位置，並提供骨科健康教育 |
| 主要資料來源 | 教育部或學術單位建置之標準化 3D 人體骨骼模型 | 商業化高精細 3D 模型與內建教材、測驗與標註資源 | 使用者上傳之骨科 X 光影像、骨骼結構資料與本研究整理之文字教材 |
| 影像／模型類型 | 可旋轉、分層的 3D 骨骼與器官模型 | 多層次、高精細 3D 解剖模型，支援多視角與多系統 | 2D X 光影像為主，未來可與 3D／混合實境應用整合 |
| 是否結合實際病人影像 | 多數不處理個人影像，著重標準模型 | 可能搭配少量示意影像，但核心仍為標準 3D 模型 | 以實際 X 光影像為起點，對應個別骨頭名稱與位置 |
| 深度學習影像偵測 | 一般未整合深度學習模型 | 一般未直接整合影像偵測模型 | 由 OBHES-IDM 採用 YOLO 進行骨骼區域偵測與標註 |
| 大語言模型與 RAG | 多數未結合大型語言模型，文字說明多為預先撰寫 | 提供固定或半結構化文字說明，較少結合 LLM 或 RAG | 由 OBHES-KRM 與 OBHES-IEDM 結合大型語言模型與 RAG，依問題與骨頭動態擷取資料並生成解說 |
| 互動方式 | 以滑鼠或觸控操作 3D 模型旋轉、放大、顯示／隱藏結構 | 提供 3D 模型操作、註記、測驗與簡易標註等多種互動工具 | 上傳 X 光、點選偵測框、以自然語言提問並與 AI 對話 |
| 個人化程度 | 以「標準人體」為主，個人化程度較低 | 主要為標準解剖結構，個人化多限於學習路徑與標註檔 | 以「個人 X 光」為核心，針對特定影像上的骨頭提供對應解說與問答 |
| 系統定位 | 國內醫學與解剖學課程之輔助教材，強調空間結構理解 | 國際醫學教育與臨床訓練常用之 3D 解剖學平台 | 醫病溝通與骨科健康教育之輔助工具，可同時支援教學與自學情境 |
| 主要優勢 | 符合本地課綱與語言環境，易於整合教學 | 模型精細度高、功能完整，具成熟生態系與資源 | 能直接對應個人 X 光影像，結合深度學習與 LLM＋RAG，自動產生可讀性較高之文字解說與互動問答 |
| 主要限制 | 難以反映個別病人的實際影像與病灶，與臨床 X 光之連結有限 | 著重標準模型與專業學習，較少納入民眾端醫病溝通與個人影像解釋 | 需蒐集並標註足夠 X 光資料與教材內容，目前雛型系統仍在架構與流程驗證階段，尚未完成大規模評估 |

綜合而言，教育部 3D 骨骼教學系統與國際 3D 解剖學平台（如：Complete Anatomy）在標準化 3D 模型與空間結構教學上具有明顯優勢，適合作為解剖學與骨骼構造課程的基礎工具；本研究提出之 OBHES 則以實際 X 光影像為起點，結合深度學習影像偵測與大型語言模型加上 RAG，提供個人化的文字解說與問答互動。三者在功能上並非互斥，而是形成互補關係：前兩者偏重「標準結構的立體理解」，OBHES 則著重「個別影像的語意解釋與溝通輔助」，共同回應骨骼教學與骨科健康教育在不同層面的需求。

# 第 3 章　系統分析與設計

### 3.1 使用情境與系統需求

本研究所建置之骨骼健康教育系統（OBHES）主要面向兩類情境：一般民眾與教學現場。對一般民眾而言，系統提供看診後再次查看骨科 X 光影像的介面，讓使用者在影像上看到各骨頭的大致位置與名稱，並透過文字提問獲得淺顯易懂的解說。對教師與學生而言，系統可作為課堂示範或自學輔助工具，利用實際 X 光影像輔助說明解剖結構，連結抽象的骨骼名稱與真實影像。

依上述情境，本研究歸納出幾項核心系統需求。首先，系統需能管理 X 光影像上傳與儲存，並透過影像標記模組支援專業人員進行骨骼位置與類別的標註與修正。其次，影像辨識模組需利用物件偵測模型自動找出影像中的骨骼區域，回傳偵測框、骨頭編號與可信度，供前端在影像上疊加框線與名稱。再者，知識檢索模組需根據骨頭編號與使用者問題，自骨骼資料與教材中擷取相關內容；聊天互動模組則將檢索結果與歷史對話整合，交由大型語言模型生成一般使用者可理解的文字回應，並紀錄對話內容以支援後續分析與系統調整。

### 3.2 系統整體架構

圖1 彙整了 OBHES 的整體三層式架構：最外層為前端使用者介面，提供 X 光影像上傳、偵測結果展示與對話視窗；中間層為應用層，由影像標記、影像辨識（OBHES-IDM）、知識檢索（OBHES-KRM）與聊天互動（OBHES-IEDM）四個核心模組構成；最內層則為資料儲存層，包含骨骼結構資料庫、影像案例與偵測結果資料表、教材表與對話紀錄，並預留未來導入獨立向量資料庫的空間。圖中可見，使用者上傳影像後，資料先經 OBHES-IDM 產生偵測框與骨骼編號，再由 OBHES-KRM 擷取相關骨骼資料與教材段落，最後交由 OBHES-IEDM 整合為自然語言回應，完成從「看圖」到「說明」的完整流程。

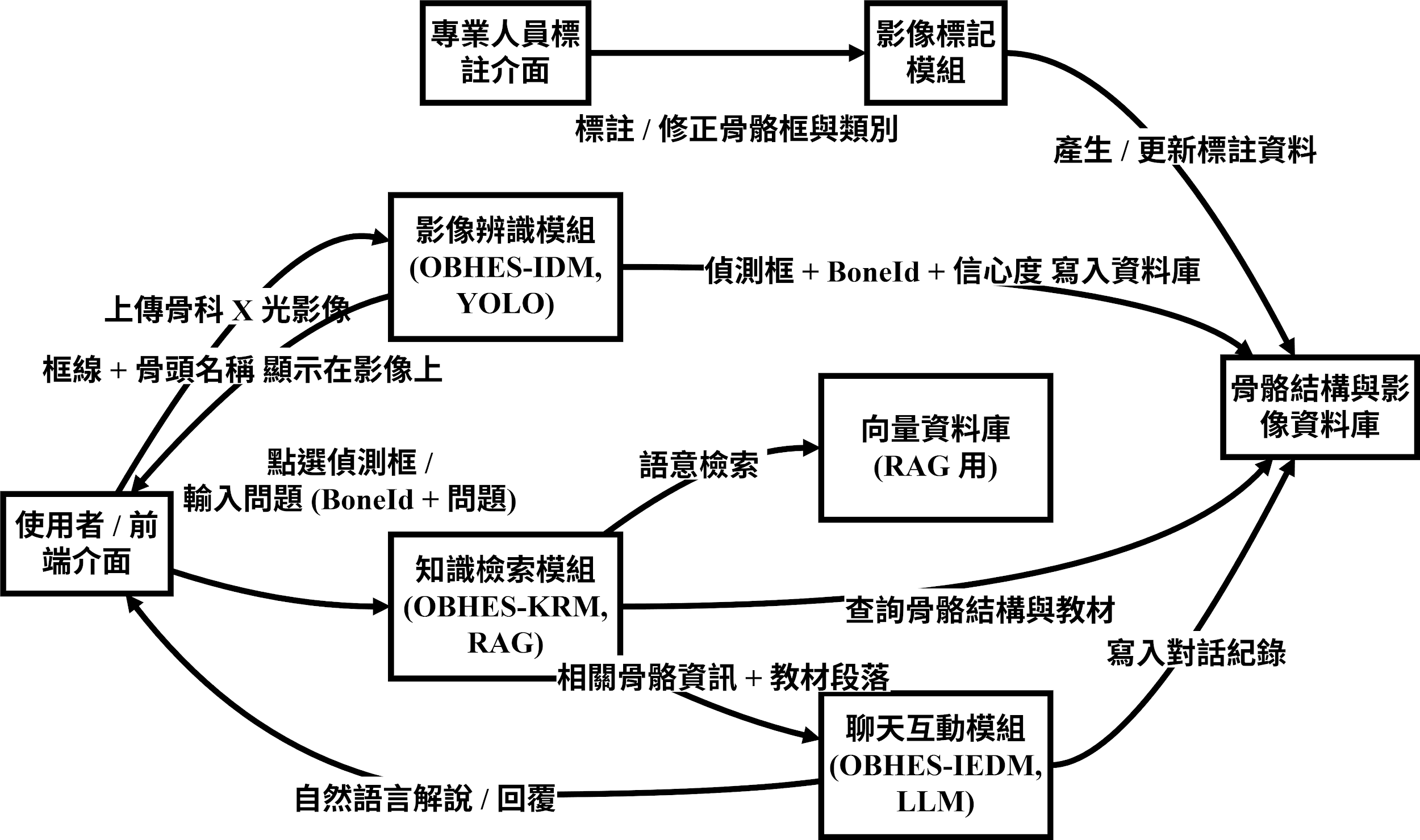


圖 1 系統架構圖

3.3 資料庫設計概要

圖2 顯示本研究後端關聯式資料庫的主要資料表以及其之間的關聯。骨骼主表負責管理骨骼大類編號與中英文名稱，骨骼細項表則補充左右別與指骨分節等資訊；兩者透過外鍵關聯，形成完整的骨骼結構層級。影像案例表記錄每一次上傳的 X 光影像與來源資訊，偵測結果表則以多筆紀錄對應同一案例，保存每個偵測框的座標、骨頭編號與信心度。教材表與對話紀錄表分別管理解剖與健康教育教材的段落，以及使用者與 AI 助理的歷史對話內容。透過圖2 所示的關聯設計，系統得以在影像偵測、骨骼結構與教材內容之間建立一致且可檢索的資料連結，支援知識檢索與聊天互動的運作。

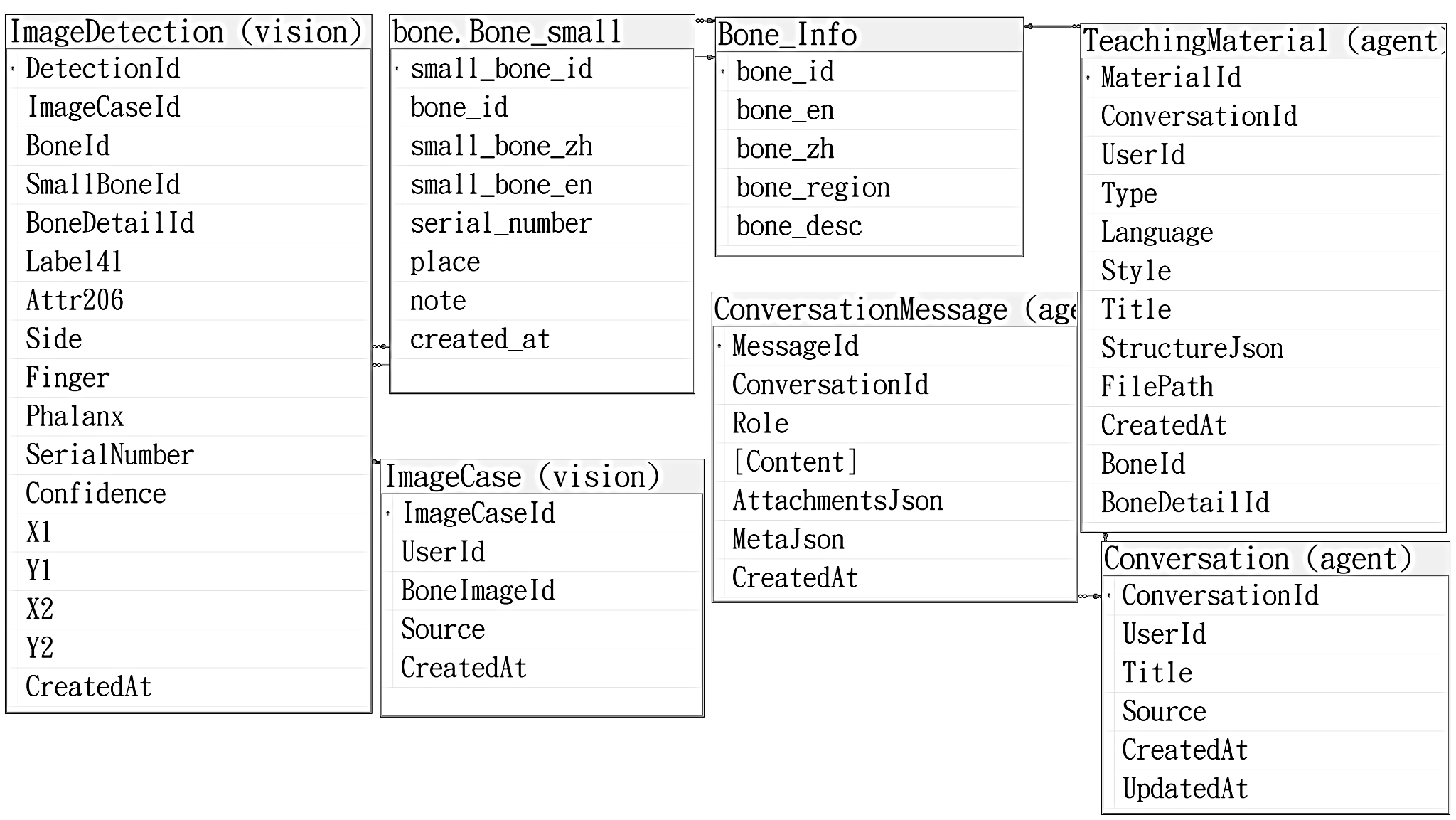


圖 2 本研究後端關聯式資料庫主要資料表和關聯示意圖

# 第 4 章　系統實作

本章說明本研究系統在實作層面的具體做法。後端以 Python 為主要語言，採用 FastAPI 作為 Web 服務框架，搭配 PyTorch 載入與執行 You Only Look Once 物件偵測模型（YOLO），實作 OBHES 影像偵測模組（OBHES-IDM）。前端採用 React 建置單頁式應用程式，負責影像顯示、框線繪製與對話介面呈現，對應 OBHES 解說與對話模組（OBHES-IEDM）的互動需求。資料儲存方面，以關聯式資料庫實作骨骼結構與對話相關資料表，用以存放骨骼基本資訊、影像案例、偵測結果與解說文字。在架構設計上，本研究預計導入獨立向量資料庫與檢索增強生成（RAG）流程；在目前的雛形實作中，則先由後端程式直接自資料庫與教材表中選取適當內容提供給大型語言模型使用，向量化與相似度檢索機制則列為後續擴充項目。

圖3 描述影像偵測模組的實作流程，從資料準備、模型訓練到推論服務建立依序串接。首先，系統蒐集骨科 X 光影像並針對選定骨骼類別進行框線標註，整理為 YOLOv8-OBB 所需的訓練格式與類別設定；接著，在可用硬體資源與資料量條件下訓練出能辨識骨骼大類的模型。完成訓練後， 影像偵測模組於後端服務中實作影像上傳端點，將前端傳入的影像進行必要前處理後交由 YOLO 模型推論，產生偵測框與類別編號，並與骨骼資料表對應，最後將偵測結果寫入資料庫並以 JSON 格式回傳前端。

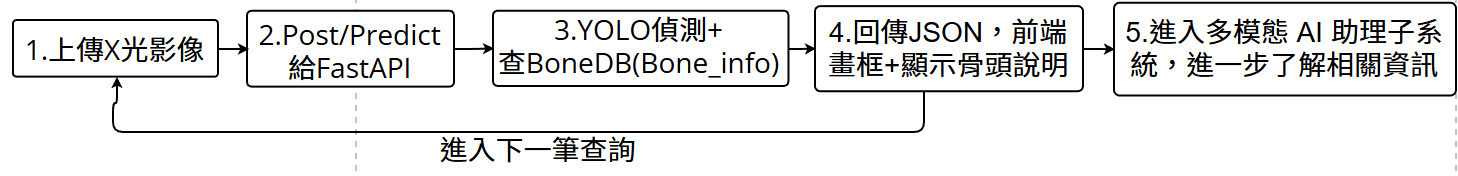


圖 3 OBHES 影像偵測模組（OBHES-IDM）實作流程圖」

圖4 說明 OBHES 智慧解說與 RAG 模組（OBHES-IEDM）的資料流。當使用者點選偵測到的骨頭或在對話視窗輸入問題時，系統會將骨骼編號與問題一併送入 OBHES-KRM，由其自骨骼結構表與教材表中擷取與該骨頭與主題相關的結構化資料與文字段落，形成「檢索結果集合」。之後，OBHES-IEDM 會將使用者當前問題、歷史對話脈絡與檢索結果整合為提示詞，交由大型語言模型生成回應，並將回應與問題存入對話紀錄表。圖4 因此展示了類似 RAG 的實作方式，即便目前向量資料庫仍為預留設計，整體流程仍已驗證從檢索到生成的可行性。

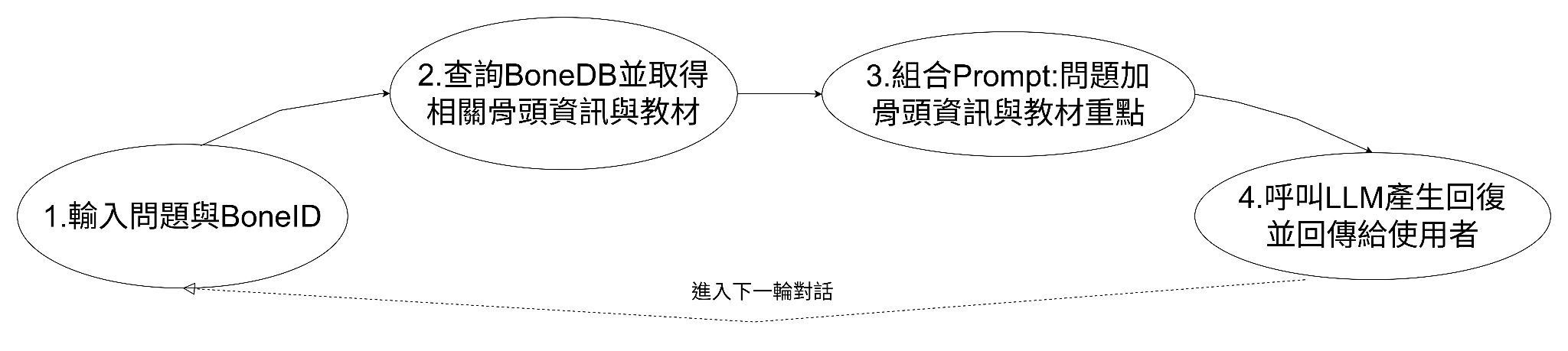


圖 4 OBHES 智慧解說與 RAG 模組（OBHES-IEM）實作流程圖」

圖5 顯示 OBHES 前端在完成 YOLO 偵測後的畫面。左側影像檢視區提供 X 光影像上傳功能，伺服器推論完成後即在影像上疊加偵測框與骨頭名稱標籤，使用者可以直覺地看到每一個骨骼區域的定位與名稱。右側則顯示偵測結果清單與基本資訊，作為後續點選骨頭與發問起點，使使用者能在同一畫面中同時掌握影像與文字資訊。

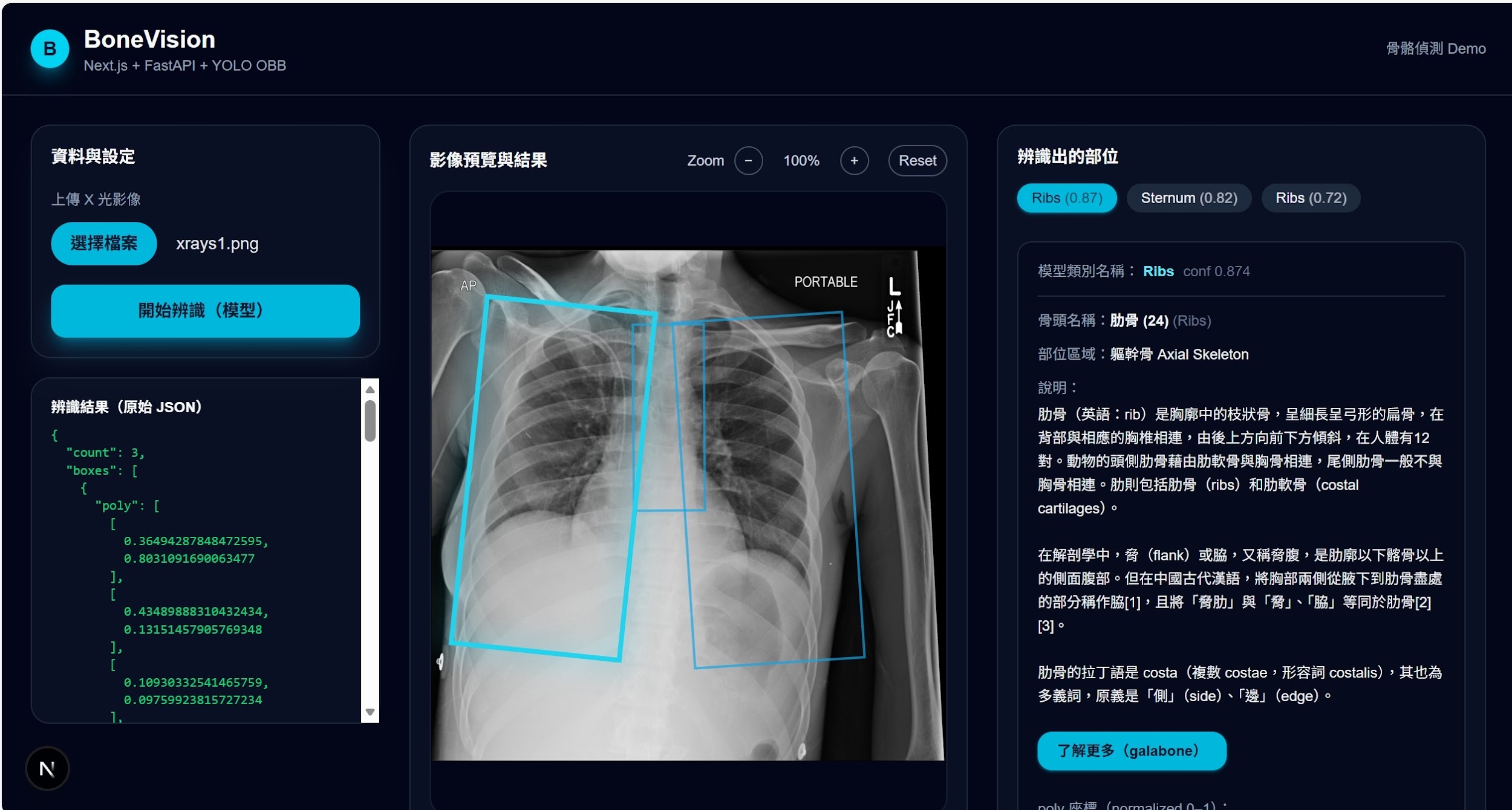


圖 5 前端畫面（影像上傳＋YOLO 偵測結果）

圖6 展示使用者在點選特定骨頭後與 AI 助理進行對話的情境。影像檢視區維持原本的偵測框顯示，當使用者點擊某一框線時，系統會在對話輸入欄位自動帶入對應骨頭名稱或編號，協助使用者聚焦於指定骨骼發問。對話區以訊息列表呈現雙方對話，AI 回應內容結合骨骼結構資料與教材段落，提供一般民眾較易理解的文字解說，讓使用者在實際影像上逐步建立對各骨骼位置與功能的認識。



圖 6 前端畫面（點骨頭＋AI 對話）

表1 將 OBHES 影像偵測模組所採用的 YOLOv8-OBB 推論流程整理為演算法。整體流程可分為四個階段：首先是影像前處理，包含將輸入 X 光影像縮放至模型要求尺寸並標準化像素值；接著由預先訓練好的 YOLOv8-OBB 模型產生候選旋轉框及其信心值；第三步透過信心值門檻與非極大值抑制（Non-Maximum Suppression, NMS），過濾掉低信心與高度重疊的框，只保留最具代表性的骨骼區域；最後根據模型輸出的類別標籤查詢骨骼資料庫，取得骨頭編號、中英文名稱與部位描述，並計算整體偵測數量。透過此演算法描述，讀者可以更清楚理解 OBHES-IDM 在程式實作層面如何由單張 X 光影像產生結構化的骨骼偵測結果。

表 1 YOLOv8-OBB 訓練演算法

|  |  |
| --- | --- |
| **演算法 1：X 光骨骼影像偵測流程（YOLOv8-OBB）** 輸入：  * image ︰ 單張 X 光影像。 * Pretrained YOLOv8-OBB Model (M)：預先訓練好的 YOLOv8-OBB 模型 。 * ：信心值門檻 (confidence threshold)。 * ：非極大值抑制 (Non-Maximum Suppression, NMS) 門檻。   流程：   1. 影像前處理 1.1 將輸入影像 I 縮放至模型要求之輸入尺寸（1024×1024）。 1.2 將像素值正規化至 [0,1]（此步驟由 YOLOv8 推論流程自動處理）。 2. YOLO 模型推論  2.1 將前處理後影像輸入模型 M。  2.2 模型輸出所有候選旋轉框 (Oriented Bounding Boxes, OBB)：   ，以及對應類別編號與信心值。   1. 依信心值過濾候選框 3.1 僅保留滿足 ≥ 之框： = 。 2. 依信心值排序 4.1 將 內所有框依信心值由大到小排序。 3. 非極大值抑制 (NMS)  5.1 依排序結果由上位框開始，逐一比較任兩框 與 之 ：  = 相交面積 / 聯集面積。  5.2 若 ，則移除信心值較低者。  5.3 最終僅保留不高度重疊且最具代表性的骨骼區域框。 4. 骨骼資訊擴充（資料庫查詢）  6.1 根據模型輸出的骨骼英文名稱，查詢骨骼資料庫，取得：  = {bone\_id, 英文名, 中文名, 部位, 描述}。 5. 輸出結果  7.1 計算最終保留之骨骼區域數量 。  7.2 輸出所有旋轉框 、多邊形座標、類別標籤、  信心值集合 ，以及對應之骨骼資訊 。   輸出：   * 偵測到之旋轉邊界框集合 。 * 各骨骼區域之類別標籤 * 信心值集合 。 * 偵測到的骨骼區域數量 。 * 由資料庫查詢得到的骨骼資訊 。 |  |
|  |  |

# 第 5 章　評估規劃與初步觀察

## 5.1 評估構想與初步觀察

本章說明 OBHES 的評估構想與目前雛型階段的實務觀察。由於本研究仍受資料量與開發時間限制，尚未完成大規模量化實驗與正式使用者研究，因此目前評估重點在於後續可執行的指標設計，以及開發過程中對系統行為的初步觀察。

在系統效能面向，未來可針對 OBHES-IDM 進行標準化的偵測效能評估，例如以 Precision、Recall、F1-score、mAP@0.5 與 mAP@0.5︰0.95 等指標，針對不同骨骼類別與影像來源比較模型表現，並檢驗在實務情境下是否需要依照信心度調整偵測門檻。聊天與解說品質方面，則可透過專家審閱與問卷調查，評估解說內容在正確性、完整性與易理解性上的表現，以及使用者對回應風格與語氣的主觀感受。

在使用者經驗面向，未來系統穩定後，可邀請具不同背景的受測者（例如資訊相關學生、醫護相關學生與一般民眾）實際操作系統，請其完成上傳影像、查看偵測結果、點選骨頭並向 AI 助理提問等任務，再透過問卷評估介面易用性與對骨骼構造理解程度。此外，亦可設計簡單的前後測題目，比較使用系統前後在特定骨骼名稱與位置辨識上的差異，以初步了解本系統作為健康教育工具的效果。

在目前雛型階段，本研究僅於開發團隊內部進行少量非正式試用，主要用於確認各模組串接是否正常、偵測結果與骨骼名稱是否對應正確，以及聊天回應是否能合理引用教材內容。這些觀察雖無統計意義，但有助於修正明顯錯誤與調整系統流程，為後續正式評估奠定基礎。

## 5.2 系統限制與風險

本研究採用實際骨科 X 光影像作為輸入，資料蒐集與標註需考量隱私與倫理議題，這也限制了可用資料的數量與多樣性。在尚未能取得大量、多來源的標註資料之前，模型的泛化能力與準確度仍具不確定性，因此本研究於目前階段刻意避免將 OBHES 定位為診斷工具，而是明確界定為健康教育與教學輔助用途。此外，大語言模型在生成自然語言說明時，仍可能出現與事實不符或過度簡化的內容。雖然本研究透過骨骼資料庫與教材內容對回應加以約束，並採用類似 RAG 的流程降低幻覺風險，仍建議在未來正式導入教學或臨床場域前，需搭配專家審查與內容驗證機制。整體而言，OBHES 目前仍屬概念與雛型階段，其主要貢獻在於系統架構與流程設計，後續仍需透過更完整的資料與實證研究，檢驗其在實際場域中的效益與限制。

圖7 以 Precision–Recall 曲線呈現 YOLOv8-OBB 模型在多類別骨骼偵測任務中的整體表現。圖中各條曲線代表不同骨骼類別在各種信心度閾值下的 Precision 與 Recall 變化，粗線則為所有類別的平均結果。由圖可見，多數類別在相當範圍內同時維持較高的 Precision 與 Recall，整體 mAP@0.5 約為 0.916，顯示模型已具備相當程度的骨骼定位與分類能力，足以支援 OBHES 在骨科健康教育情境下標示「這是哪一塊骨頭」的需求。

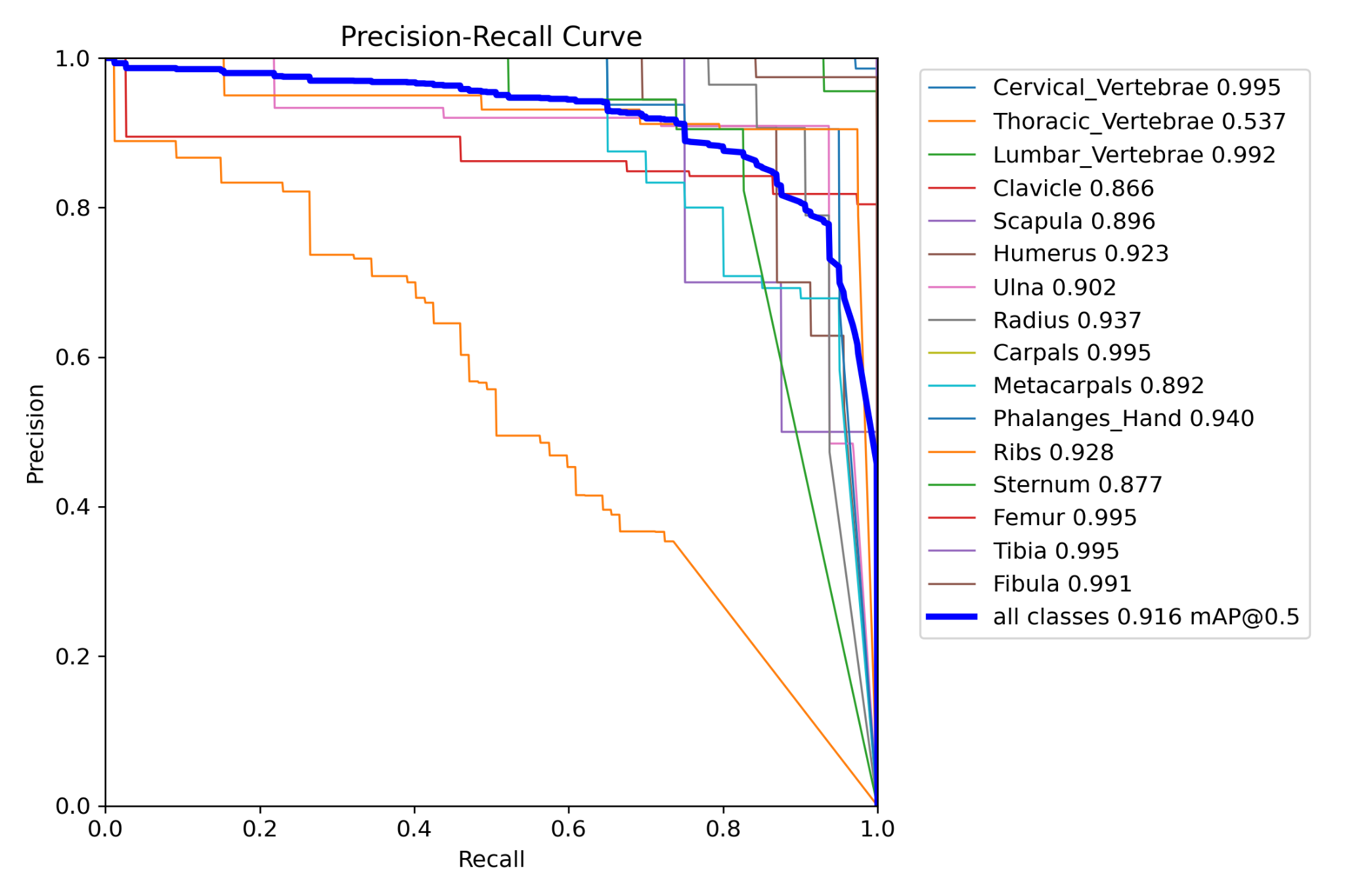


圖 7：Precision–Recall Curve（PR 曲線）

圖8 顯示模型在各骨骼類別上的標準化混淆矩陣，深色對角線代表正確辨識比例，非對角線的格則代表可能的誤判類別。大部分骨骼類別皆呈現較高的對角線值，顯示模型能有效區分多種骨骼結構；但在解剖位置接近或影像特徵相似的類別，例如頸椎與胸椎、鎖骨與肩胛骨，仍可觀察到一定程度的混淆。這樣的結果一方面說明骨科影像辨識的固有困難，另一方面也提供後續資料增強與模型調整的方向，作為未來提升 OBHES 影像偵測穩定性的依據。

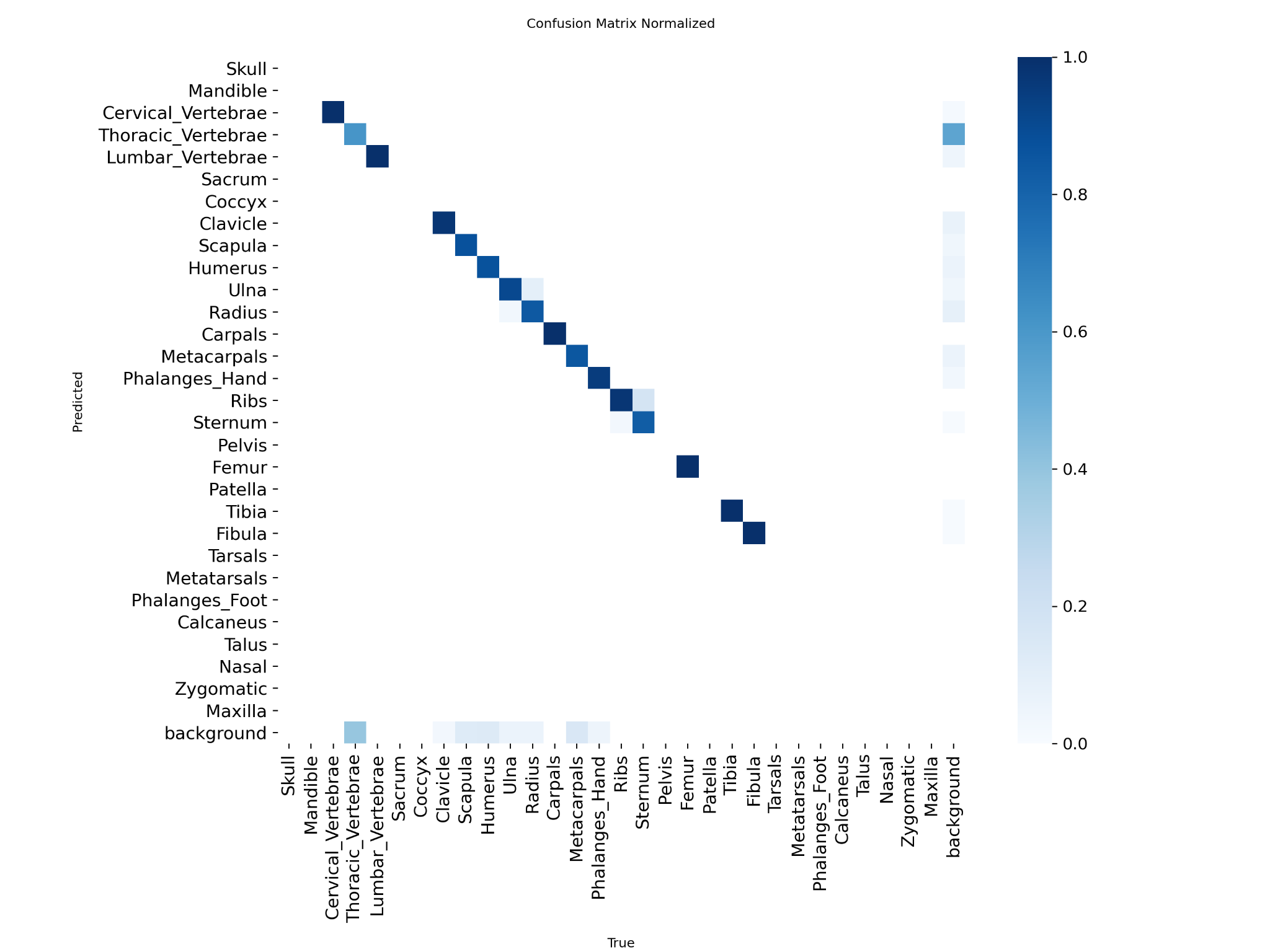


圖 8：Confusion Matrix（Normalized）

## 5.3 使用者互動與健康教育成效評估規畫

在使用者互動與健康教育成效方面，本研究規畫以問卷與操作任務作為主要評估工具。未來系統穩定後，可邀請具不同背景的受測者（例如資訊相關學生、醫學或護理相關學生、一般民眾）實際操作系統，請他們完成上傳影像、查看偵測結果、點選骨頭並向 AI 助理提問等任務，再透過問卷收集對介面易用性、解說文字可理解性、對骨骼構造理解程度等面向的主觀評價。

此外，亦可設計簡單的前測與後測題目，評估使用者在使用系統前後對特定骨骼名稱與位置的掌握情形，以初步了解系統作為健康教育工具的效果。在目前雛形階段，本研究僅在開發團隊內部進行少量非正式試用，主要用於確認功能是否按預期運作，尚未進行具有統計意義的正式研究。

# 第 6 章　結論與未來研究方向

本研究提出一套結合 YOLO 影像偵測與大型語言模型之骨科醫療影像健康教育系統，稱為「骨骼健康教育系統」（OBHES）。系統整合影像標記、影像辨識（OBHES-IDM）、知識檢索（OBHES-KRM）與聊天互動（OBHES-IEDM）四個核心模組，讓使用者能上傳骨科 X 光影像、查看骨骼偵測結果，並透過自然語言向 AI 助理詢問與特定骨頭相關的解說內容。

在技術實作上，本研究蒐集並標註胸廓區域 X 光影像，使用 YOLOv8-OBB 模型進行骨骼偵測，並設計後端資料庫管理骨骼結構、影像案例、偵測結果與教材段落，同時預留未來導入向量資料庫與完整 RAG 流程的空間。初步結果顯示，雖然資料量有限，模型在目標骨骼類別上的 mAP@0.5 已達約 0.916，足以支援雛型系統中「定位骨骼並連結知識內容」的需求。

本研究的主要貢獻在於：一、提出以「使用者個人 X 光影像」為中心的骨科健康教育系統架構，補足現有 3D 教學平台多以標準模型為主、較少連結實際影像的缺口；二、示範如何將深度學習影像偵測、骨骼結構資料庫與 LLM＋RAG 概念整合於同一系統流程之中；三、在系統開發層面，建立可持續擴充的標註、訓練與知識整理機制，為後續研究與實務應用提供基礎。

未來研究可在數個方向延伸。資料面向上，可擴充影像來源與骨骼類別，納入不同年齡、不同部位與多中心資料，並持續優化偵測模型。在系統面向上，可導入獨立的向量資料庫與自動化 RAG 管線，提升知識檢索的彈性與可維護性。互動面向則可針對不同使用族群（例如民眾、醫學生、護理人員）調整解說深度與語氣，並與現有 3D 解剖學平台或混合實境應用整合，搭配系統化的使用者研究，以實證方式檢驗 OBHES 在實際教學與醫病溝通場域中的效益與限制。

參考文獻

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Bakhtin, A., … Zampetakis, M. (2023). GPT-4 technical report. arXiv preprint arXiv︰2303.08774.

Alzubaidi, L., Al-Dulaimi, K., Salhi, A., Alammar, Z., Fadhel, M. A., Albahri, M. F., … Gu, Y. (2024). Comprehensive review of deep learning in orthopaedics︰ Applications, challenges, trustworthiness, and fusion. Artificial Intelligence in Medicine, 155, 102935.

Ju, R. Y., & Cai, W. (2023). Fracture detection in pediatric wrist trauma X-ray images using YOLOv8 algorithm. Scientific Reports, 13, 20077.

Lee, J., & Chung, S. W. (2022). Deep learning for orthopedic disease based on medical image analysis︰ Present and future. Applied Sciences, 12(2), 681.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., … Riedel, S. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems, 33, 9459–9474.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., … Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60–88.

Motsinger SK. Complete Anatomy. J Med Libr Assoc. 2020 Jan;108(1)︰155–7. doi︰ 10.5195/jmla.2019.853. Epub 2020 Jan 1. PMCID︰ PMC6920010. https︰//pmc.ncbi.nlm.nih.gov/articles/PMC6920010/.

Srinivasu, P. N., Kumari, G. L. A., Narahari, S. C., Ahmed, S., & Alhumam, A. (2025). Exploring the impact of hyperparameter and data augmentation in YOLO V10 for accurate bone fracture detection from X-ray images. Scientific Reports, 15, 9828.

Su, Z., Adam, A., Nasrudin, M. F., Ayob, M., & Punganan, G. (2023). Skeletal fracture detection with deep learning︰ A comprehensive review. Diagnostics, 13(20), 3245.

Alzubaidi, L., Al-Dulaimi, K., Salhi, A., Alammar, Z., Fadhel, M. A., Albahri, M. F., … Gu, Y. (2024). Comprehensive review of deep learning in orthopaedics︰ Applications, challenges, trustworthiness, and fusion. Artificial Intelligence in Medicine, 155, 102935.