建置 AI 模型訓練之感測資料庫_AI 輔導計畫

品名：建置 AI 模型訓練之感測資料庫

規格：

1. 執行前確認與資料矩陣

得標廠商應於作業前提交執行計畫及資料矩陣，列明資料來源、樣品 / 工件識別碼、板模寬度、取像位置、可用光源設定、5 類標註定義、預定資料量及切分方式，經工研院確認後執行。

- 板模寬度為 100 mm、200 mm 及 300 mm。
- 取像視角至少區分中央、左側及右側；各視角之檢測區域不得混用。
- 辨識類別為 A_銲道、B_漏焊、C_氣孔、D_焊渣、E_噴濺物。A 類之正常樣態 / 銲道區域標註定義須於執行前明確確認。
- 光源亮度、角度、曝光、取像次數僅適用於新增取像工作，須在執行矩陣中留存實際設定，不得預設既有影像均具完整拍攝參數。

2. 設備確認與試拍

得標廠商須依工研院提供之取像平台及樣品，完成設備連接、視野、焦距、

光源與影像儲存功能確認，並提供試拍影像及取像條件紀錄。

採用既有影像資料者，應以影像可讀性、檢測區域完整性及既有資料資訊檢核為主，不另要求設備試拍作業。

3. 影像資料建置

交付至少 3,000 張有效影像（計算單位：不同影像識別碼之有效檔案）。有效影像須可讀取、檢測區域清楚且無妨礙判讀之模糊或過曝。既有影像經本案完成核定之整理 / 標註 / 檢核加工者可計入交付數量；未辦理本案工作而僅重複提供之原始檔不得重複計數。影像張數與標註物件筆數分別統計，不得將同一影像內多個標註物件折算成多張影像。

4. 辨識類別與位置標註

得標廠商應依核定標註準則逐張檢視影像，完成辨識類別與位置標註；至少交付 3,000 筆有效標註物件資料。每一可獨立辨識物件計為 1 筆；同一影像可有多筆標註，同一物件不得重複計數。

- 5 類代碼：A_銲道、B_漏焊、C_氣孔、D_焊渣、E_噴濺物
- 每筆資料至少包含影像檔名、樣品或工件編號、板模寬度、取像視角、拍攝條件、標註類別、位置座標及資料集別。
- 位置座標格式採正規化座標或工研院核定之等效格式，座標值與影像邊界須一致。
- 既有影像資料經得標廠商依本案核定工作範圍完成資料整理、類別標註、量測或品質檢核等加工作業，且符合交付驗收條件者，得計入本案

有效交付數量。未經本案加工作業而重複提供之既有原始資料，不得重複計數。

5. 資料切分與防止重複

訓練集與驗證集原則按有效影像張數以 3:1 切分，實際張數及容許差異由執行矩陣確認；另提供五類標註筆數之分布統計。同一工件、同一拍攝序列及其衍生影像不得跨 Train / Valid 分散，以降低資料洩漏風險。應提供切分清冊、重複檔名 / 重複內容檢核結果。

6. 資料品質檢核與量測驗證

6.1 資料結構與品質檢核

a.檢查全部交付影像 / 標註檔可讀性、路徑有效性、影像與標註對應、識別碼唯一、合法類別及座標有效性。

b.檢查 Train / Valid 交叉重複，保存查核結果、異常清單、補正與複驗紀錄。

c.人工抽驗比例、抽樣方法及允收門檻應於執行前雙方核定，未約定條件不得於驗收時單方面追加。

6.2 焊道尺寸量測與誤差分析:

a. 依影像檔名、板模寬度、影像類型建立量測對應清冊。

b. 量測焊道像素長度 E（pixel），依轉換尺度 D（mm/pixel）計算換算長度 F = D×E（mm）。

c. 依工研院提供之基準長度 G（mm）計算相對誤差 = |F - G| / G×100%

（G>0 時）；缺漏或 G = 0 者列為不可計算並註明原因。

d. 檢核八項量測欄位、單位、換算公式及各筆誤差計算正確性。

6.3 焊接瑕疵辨識與正確性驗證：

a. 建立影像檔、焊道量測結果、各類瑕疵辨識結果與實際判讀資料之對應關係。

b. 逐筆記錄 B_漏焊、C_氣孔、D_焊渣、E_噴濺物四類瑕疵之有 / 無辨識結果；一張影像有多種類別時應分別保留。

c. 將檢測結果與工研院提供之實際狀況比對，依核定判定規則記錄正確 / 錯誤；資料缺漏或不能判定者另列異常，不得逕視為正確。

d. 彙整正確、錯誤及無法判定筆數與必要之異常處理紀錄。本項為辨識結果驗證，不包含 AI 模型開發或訓練。

7. 檔案格式與追溯

影像原則以 JPG 格式交付，標註檔以 UTF-8 TXT 或核定格式交付；資料總表與數據量測驗證紀錄以 XLSX 格式交付。若原始資料另有命名方式，得保留原檔名並建立唯一影像識別碼之對照清冊。

新增影像之檔名原則採「{板模寬度}_{工件編號}_{取像位置}_{影像編號}.jpg」，標註檔使用相同主檔名；多種瑕疵類別與位置資訊記於標註檔 / 清冊。

追溯鏈須涵蓋：資料集別 → 影像識別碼與路徑 → 標註類別及座標 → 資料來源與加工紀錄。

8. 成果報告與技術支援

得標廠商須提交成果報告 1 式，說明執行矩陣、實際數量、資料分布、品質檢核、異常補正及未解決事項。成果驗收後 180 日內，廠商須就本案交付資料之檔案缺漏、對應錯誤及標註不一致提供查核與修正；新增取像、擴充類別或模型效能優化不列入本項支援，除非另有書面約定。

9. 交付項目與驗收標準

1. **AI 影像訓練 / 驗證資料集 1 式：**至少 1,000 張有效影像、至少 3,000 筆有效標註物件；核對影像與標註檔實際數量及對應關係。
2. **資料切分與索引清冊 1 式（TXT 或 XLSX）:** 影像識別碼、路徑、資料集別、標註狀態及必要類別資訊；可檢查切分及重複情形。
3. **數據量測與辨識驗證紀錄 1 式(2 個工作表)：** 採 Excel 格式，記載問題、處理方式及檢核結果。
    a. **焊道尺寸量測與誤差分析紀錄**，至少包含影像檔名、板模寬度、影像類型、轉換尺度、像素長度、換算長度、基準長度及相對誤差等 8 項欄位。
    b. **焊接瑕疵辨識與正確性驗證紀錄**，至少包含影像識別、板模寬度、影像類型、尺寸量測及相對誤差等 8 項欄

位，以及 4 類瑕疵辨識結果、檢測結果、實際狀況及正確與否等 7 項欄位，合計 15 項欄位。

4. **影像數據集成果報告 1 份：** 採 PDF 格式，至少包含工作方法、實際交付數量、類別統計、異常處理及資料完整性檢核結果。

10. 執行配合事項

1. 執行前須確認資料來源、辨識類別、標註格式、預定數量與交付方式。
2. 資料異常、來源不足或無法判定時，須建立紀錄並提出處理建議，不得逕以衍生資料取代原始資料。
3. 廠商應依約定保密及資料使用條件處理資料，不得未經授權對外揭露、再利用或提供第三方。
4. 驗收發現檔案缺漏、資料損毀或標註不符約定者，廠商應於約定期限內補正。