國立台北大學資訊工程學系專題報告
基於超聲波之蝙蝠物種辨識系統

專題組員:曾柏碩、林子欽、蔡承祐、陳立亞

專題編號:PRJ-NTPUCSIE-114-14
執行期間:2025年9月 至2026年6月

1.摘要

蝙蝠在台灣生態系中占有重要地位與保育價值。本專題旨在建立一套基於超聲波音訊之蝙蝠物種自動辨識系統，以解決傳統聲學監測方法高度依賴人工判讀、耗時且難以擴展之問題。系統以蝙蝠回聲定位音訊為主要分析對象，整合音訊前處理、事件偵測、雜訊篩選與深度學習分類等流程，實現自動化物種辨識。

本系統於前處理階段先以頻譜減法抑制背景雜訊，並結合高頻濾波強化目標訊號。接著透過滑動視窗切分音訊，並以小波轉換進行多尺度事件偵測，以篩選可能含蝙蝠叫聲之片段，並進一步使用 Random Forest 進行二階段雜訊過濾，以提升資料品質。

在特徵建構方面，採雙分支架構：其一將聲譜圖轉為三通道影像，輸入 Vision Transformer 以擷取時頻特徵；其二使用 IMFCC 搭配 1D-CNN 建模高頻細節。最後融合兩維度特徵進行物種分類。

實驗結果顯示，本系統在獨立測試集上達到 95.10% 準確率，在盲測資料上亦達 84.08% 之辨識表現，驗證所提出方法於蝙蝠聲學辨識任務之可行性與穩定性。未來可透過擴充資料集與改善雜訊環境適應能力，進一步提升模型於實際野外監測情境之應用效能。

2.簡介

2.1.研究背景

在生態研究與環境保育領域中，研究人員會透過直接證據（如目擊紀錄、聲音監測）以及間接證據（如排遺、足跡等），調查特定區域與棲地中的動物分布情形，藉此了解生態系統的組成與變化。

蝙蝠是哺乳動物中唯一具備主動飛行能力的類群，在生態系中扮演重要角色，包括抑制害蟲數量、協助植物授粉以及傳播種子等功能。此外，蝙蝠的物種多樣性也常被作為評估環境品質與生物多樣性的指標之一。根據目前資料，台灣已記錄到 37 種蝙蝠，約佔台灣哺乳動物種類的46%，顯示蝙蝠在台灣生態系中的重要地位與保育價值。

2.2.研究動機

蝙蝠的調查相當困難，因其具有夜行性與棲地隱密特性，目前主要的調查方式有三種:第一種是找到日間居所，直接目擊；第二種是夜間架網捕捉；第三種為高採樣率之錄音設備側錄其回聲定位叫聲，並以頻譜圖觀測。然而蝙蝠日間棲所不易找尋，侵入棲地會干擾行為，蝙蝠利用回聲定位又容易躲過捕捉網，成效不佳。

近年來，研究人員逐漸利用蝙蝠回聲定位叫聲進行物種辨識與生態調查，此方法能補足傳統捕捉方式難以記錄的物種資料。然而，蝙蝠叫聲屬超音波頻段，需透過專業設備錄製，且其頻譜圖判讀高度仰賴專業知識，目前仍以人工辨識為主。由於單次叫聲僅持續數十毫秒，大量音訊資料的分析十分耗時，使大規模蝙蝠監測與多樣性研究面臨挑戰。

目前台灣尚未有一套完整的蝙蝠音訊監測系統。因此，本專題以蝙蝠回聲定位音訊為研究對象，開發一套基於超聲波音訊之蝙蝠物種自動辨識

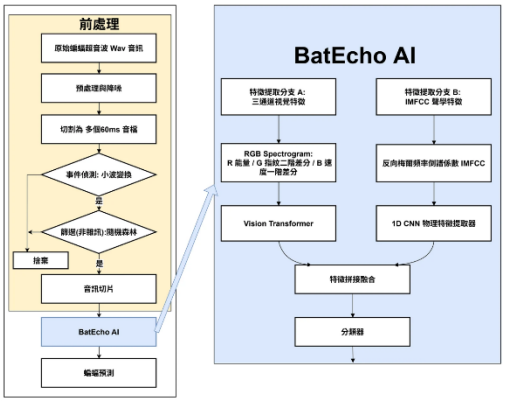
系統，期望降低聲學資料分析門檻，並提升蝙蝠生態調查的自動化程度，進而達到大規模監測之目標。

2.3. 研究目的

本專題旨在建置一套基於超聲波音訊的蝙蝠物種自動辨識系統，透過深度學習模型分析蝙蝠回聲定位訊號，協助研究人員快速完成物種分類工作。研究目標包括：

- 建立蝙蝠超聲波音訊之前處理與事件擷取流程。
- 建構深度學習模型進行蝙蝠物種分類。
- 設計使用者操作介面，提供音檔輸入與辨識結果呈現功能。
- 評估系統於實際錄音資料上的辨識效能與應用可行性。

2.4. 研究架構



【圖一 系統架構圖】

本系統透過音訊前處理、第一階段的事件偵測與第二階段的雜訊篩選，從原始錄音中擷取較具代表性的蝙蝠叫聲片段，並結合深度學習模型進行物種分類。模型設計上同時考量聲譜影像特徵與高頻聲學特徵，以提升系統對不同錄音環境與蝙蝠叫聲型態的辨識能力。

此外，本專題亦建置使用者介面，支援音檔輸入、模型版本選擇、分類結果顯示、信心指標呈現與聲譜比對等功能，使研究人員能更直觀地進行蝙蝠聲音資料分析。

3. 專題進行方式

3.1. 開發平台

本專題採用Ubuntu系統作為工作站進行開發，詳細規格如下:

作業系統(OS)：

Ubuntu 20.04.6 LTS

處理器(CPU)：

Intel(R) Xeon (R) CPU E5-2630 v4 @ 2.20GHz

顯示卡(GPU)：

NVIDIA GeForce RTX 3090

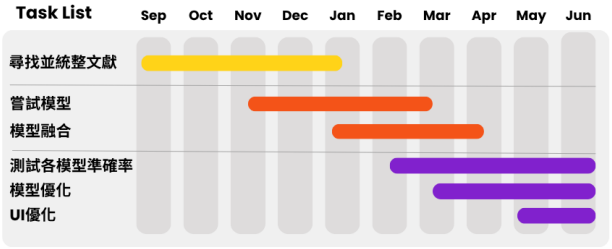
3.2. 資料集

陳湘繁學者、林清隆學者、黃俊嘉學者、張恒嘉學者，提供我們大量的蝙蝠聲音樣本。

3.3. 人員分工

曾柏碩	文獻整理、平台架設、盲測系統、模型測試、介面UI
林子欽	文獻整理、資料集整理、專案整理、模型測試、介面UI
蔡承祐	文獻整理、Focalloss處理、模型測試
陳立亞	文獻整理、前處理方法測試及整理、模型測試

3.4. 時程規劃



- **初期：** 以尋找與統整相關文獻為主，奠定理論與技術基礎。
- **中期：** 進行模型的嘗試、組合與初步測試。
- **後期：** 針對模型效能與使用者介面進行全面優化。

3.5. 系統設計

3.5.1. 前處理

在野外環境中錄製蝙蝠回聲定位音訊時，音檔常包含背景音、低頻干擾與錄音設備雜訊。若直接將原始音訊進行切片與分類，後續事件偵測與分類模型可能受到非目標聲音影響。因此，本系統於音訊輸入至模型訓練前，先對音訊進行預處理，以降低背景雜訊對後續訓練流程之干擾。

3.5.1.1. 頻譜減法與降噪

本系統先將原始音訊訊號（x(t)）透過短時傅立葉轉換(STFT)轉換至頻率域，取得其頻譜表示（X(f,t)）。接著，利用音檔前10%訊號估計背景雜訊頻譜模板（N(f)），並自原始頻譜中扣除該背景雜訊，其概念可表示為：

$$\hat{S}(f,t) = \max\left(|X(f,t)| - N(f), 0\right)$$

其中，(|X(f,t)|) 表示原始音訊於頻率（f）與時間（t）下之頻譜幅度，(N(f)) 為估計之背景雜訊頻譜，($\hat{S}(f,t)$) 則為降噪後保留之頻譜。透過此方法，可抑制音檔中的背景雜訊，並保留較可能為蝙蝠回聲定位之音訊特徵。

此外，由於蝙蝠回聲定位音訊主要位於高頻區域，我們進一步將 35 kHz 以下之低頻成分視為非蝙蝠回聲定位訊號予以移除，同時對 35–50 kHz 區間之背景干擾進行抑制。經頻譜減法處理後，音訊再進入後續切片、事件偵測與資料篩選流程，以提升分類模型之輸入資料品質。

3.5.1.2. 音訊切片

經頻譜減法降噪後，系統將完整音訊切割為固定長度之短時間片段，以便後續進行事件偵測與特徵分析。由於蝙蝠回聲定位叫聲通常具有持續時間短、頻率變化快速之特性，若直接以完整音檔來分析，容易受到長時間背景雜訊與無效區段影響。因此，本系統採用滑動視窗方式進行音訊切片，將音訊切割為長度 60 ms 之片段，並以 20 ms 作為視窗位移量。

此切片方式使相鄰片段之間保有重疊區域，可降低蝙蝠叫聲剛好出現在片段邊界而被截斷或遺漏之可能性。其切片起始位置可表示為：

$$s_i = i \times L_{shift}$$

其中，(s_i) 表示第（i）個片段之起始位置，(L_{shift}) 表示每次視窗位移長度。因此相鄰片段之間具有約 66.7% 的重疊比例。

完成切片後，每一段音訊片段並不會直接作為分類模型之輸入，而是會進一步經由小波事件偵測判斷是否包含疑似蝙蝠回聲定位事件。僅有通過事件偵測門檻之片段才會被保留，以減少無效音訊與背景雜訊片段進入後續辨識流程，提升系統整體分析效率與分類資料品質。

3.5.2. 事件偵測、篩選

若將高解析度音訊每個切片輸入後續模型運算，會是一個很龐大的運算負擔，因此我們設計了兩個階段的篩選，進一步篩選出目標的音訊。

3.5.2.1. 事件偵測

完成音訊切片後，本系統針對每一個短時間片段進行事件偵測，以判斷該片段是否可能包含蝙蝠回聲定位音訊。由於蝙蝠叫聲具有時間短暫、能量集中且頻率變化快速之特性，單純依據整段能量判斷容易受到背景雜訊或短暫干擾影響。因此，本系統採用小波轉換進行多尺度分析，以同時

觀察音訊在時間與頻率上的變化特徵。

本系統使用 Daubechies 4（db 4）小波作為基底函數，將音訊片段分解為不同尺度之小波係數。若片段中存在明顯的回聲定位事件，該尺度的小波係數會出現較高振幅變化。我們以各尺度係數之平均值與標準差建立事件判斷門檻，其概念可表示為：

$$T_j = \mu_j + k\sigma_j$$

其中，T_j 表示第 j 個尺度之事件判斷門檻，μ_j 為該尺度小波係數絕對值之平均值，σ_j 為其標準差，k 為門檻調整參數。為防止單一尺度雜訊造成誤判的可能性，本系統要求片段中至少有多個尺度同時超過門檻的變化，才判斷其屬於有事件出現，以排除整體能量過低之片段。

3.5.2.2.音訊篩選

經小波事件偵測後，系統可篩選出疑似包含蝙蝠回聲定位事件之音訊片段。然而，小波偵測主要依據短時間能量變化與多尺度係數進行判斷，部分環境雜訊、昆蟲聲或其他非目標聲學事件仍可能通過第一階段篩選。因此，本系統採用 Random Forest 作為第二階段資料篩選方法，將音訊片段區分為有效片段與雜訊片段，以降低非目標聲音進入後續分類模型之比例。

在此階段中，系統針對每一個通過小波偵測之片段萃取聲學統計特徵，包括 MFCC、頻譜中心、頻譜平坦度與頻譜滾降點等。上述特徵分別可描述聲音頻譜形狀、能量集中位置、音調或雜訊特性，以及主要能量涵蓋之頻率範圍，並作為 Random Forest 判斷片段品質之輸入。

Random Forest 由多棵決策樹組成，各決策樹根據不同特徵組合進行分類，最後透過多數決產生最終判斷，其概念可表示為：

$$\hat{y} = majority(h_1(x), h_2(x), \ldots, h_t(x))$$

其中，x 表示音訊片段之聲學特徵向量，$h_t(x)$ 表示第 T 棵決策樹之分類結果，$\hat{y}$ 則為多數決後得到之最終分類結果。本系統僅保留有效片段進入後續物種分類流程，以提升深度學習模型之輸入資料品質與辨識穩定性。

3.5.3.模型框架

本專案主要模型框架分為兩條特徵分支，分別為三通道視覺特徵(簡稱分支A)、IMFCC聲學特徵(簡稱分支B)，並透過兩條分支特徵結果結合作為本專題的最終模型。

3.5.3.1.三通道視覺特徵(分支A)

此分支的核心概念是將音訊訊號「影像化」，並賦予 RGB 三個通道不同的物理與聲學意義，讓ViT 模型能像處理圖片一樣，去萃取音訊在時間與頻率上的紋理特徵。

R 通道（能量強度 - 頻譜原貌）：
將原始 Log-Mel 頻譜圖的分貝值線性正規化至 [0, 1] 區間。象徵頻譜分貝的能量分布。

G 通道（突變輪廓 - 二階差分）：
計算頻譜矩陣的二階差分。如同影像處理中的「邊緣檢測」。強調尖銳、極短的聲音形狀，凸顯出能量瞬間突變的輪廓。

B 通道（動態趨勢 - 時間導數）：
使用一階導數（Δ）計算頻譜在時間軸上的變化率，並正規化至 [0, 1]。這個通道賦予了模型「動態」的

視角，專門捕捉音訊頻率隨時間滑動或漸變的動態行為。

3.5.3.2. IMFCC聲學特徵(分支B)

分支B從聲學頻譜角度分析蝙蝠叫聲，由於蝙蝠叫聲集中於高頻區域，因此使用IMFCC強化高頻解析度。

IMFCC

對輸入訊號進行短時傅立葉變換，計算其功率譜，功率譜通過一組在高頻區域配置更密集的濾波器組(Inverse Mel Filter Bank)，經濾波器能量加總後取Log，再進行離散餘弦變換（DCT）以解除特徵間的相關性，最終輸出 N維的 IMFCC 倒譜係數序列。

CNN

將IMFCC視為隨時間變化的聲學序列，使用1D-CNN擷取局部時間變化與高頻聲學模式，輸出聲學特徵向量，供後續與ViT特徵融合。

4. 實驗

經過前面所述的特徵萃取與融合操作後進入訓練，整體的模型驗證與評估機制規劃如下：

4.1. 模型訓練 (Training Phase)

數據集分割與防止洩漏機制:為了確保模型評估的客觀性，避免同一次錄音切出的不同片段同時出現在訓練集與測試集中，系統採取了嚴格的防洩漏機制。

4.1.1. 獨立測試集切分

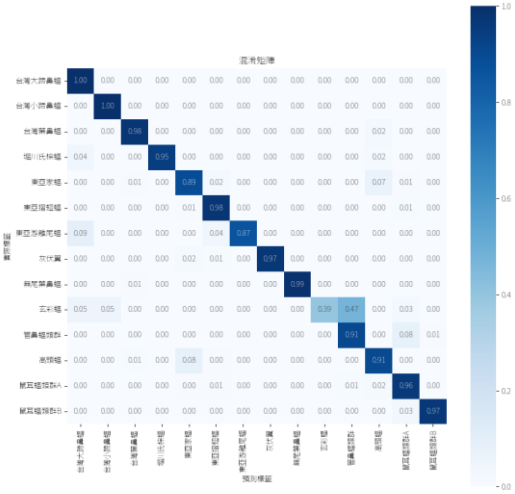
使用 GroupShuffleSplit劃分出20%的獨立測試集，以音檔來源（資料夾名稱）作為 group_id。

4.1.2. 交叉驗證

在訓練集內部，採用5-Fold StratifiedGroupKFold。確保每一個fold validation set 的音檔來源與 training set 徹底隔離，還維持了各fold間類別分佈的一致性。

4.2. 模型測試(Testing Phase)

在獨立的 20% 測試集（共計 16,659 個樣本）上，模型取得了優異的成績，整體準確率達到了 95.10%。



【圖二 混淆矩陣】

4.3. 盲測驗證 (Blind Test Phase)

各蝙蝠類別之分類效能評估結果

Class	Precision	F1 Score
台灣大蹄鼻蝠	1.0000	1.0000
台灣小蹄鼻蝠	0.5714	0.7273
台灣葉鼻蝠	1.0000	0.8750
東亞摺翅蝠	1.0000	0.4444
玄彩蝠	1.0000	1.0000
管鼻蝠類群	0.7500	0.8182
鼠耳蝠類群A	1.0000	0.4000
鼠耳蝠類群B	0.2222	0.3077
Accuracy	--	0.7174
Weighted Average	0.8408	0.7154

表格 1 盲測資料辨識結果

本專題經過盲測結果後整體平均準確度為84.08%，如表格1所示。

5. 主要困難與解決辦法

5.1. 壞軌與採樣率不一致

部分音檔損壞或因錄音設備不同導致採樣率懸殊。

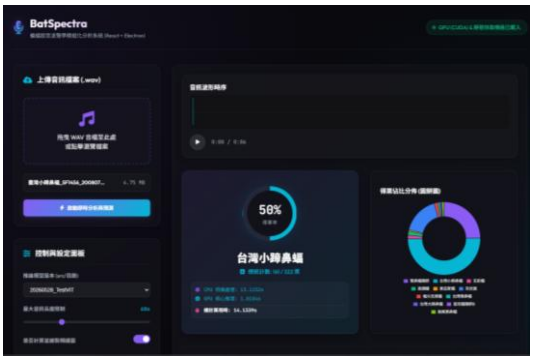
5.2. 解決辦法

透過觀察頻譜特徵還原正確採樣率，後續統一篩選並保留 384 kHz 採樣率的資料集以防干擾 。

6. 主要成果與評估

當前的蝙蝠自動辨識系統已能完成從音檔輸入到輸出蝙蝠物種辨識結果的任務，包含音訊前處理、事件偵測、Random Forest 篩選、雙分支分類模型以及使用者介面。

使用者介面能支援音檔輸入、選擇使用之模型版本、物種辨識結果以及頻譜圖對比，可協助研究人員用來進行聲學資料分析，如圖三所示。



【圖三 系統使用者介面】

於表格1中可看到，目前的實驗結果顯示整體盲測辨識準確率達 84.08%，此結果可看出本系統對應用於蝙蝠回聲定位之物種辨識具有不錯的成果，惟部分物種或類群因聲學特徵相近，或是音檔依舊受背景雜訊影響，而導致仍可能產生誤判。

7. 結語與展望

本專題成功開發出一套基於超聲波音訊之蝙蝠物種自動辨識系統，解決了傳統仰賴人工判讀且耗時的問題。系統整合了頻譜減法、滑動視窗切片與小波轉換事件偵測等前處理技術，並結合 Random Forest 進行二階段音訊過濾，大幅提升了音訊資料的有效性。

未來團隊期望能將此系統部署於全天候不間斷的監測環境。另外為了克服實際盲測的瓶頸並優化系統資源消耗，將著重於以下兩大方向：

- **資料增強與泛化能力提升**：計畫引入 Mixup 演算法等資料增強技術，提升模型面對環境雜訊干擾時的泛化強度。
- **消融實驗與模型輕量化**：為優化雜訊處理流程並精簡資源，未來將進行消融實驗，拆解並評估各模組對準確率的實際影響。藉由剔除冗餘的結構，大幅降低運算成本與訓練時間。

8. 銘謝

感謝教授的指導與學長姐的協助，引導我們專題的進行方向及順利完成專題設下的目標。

另外也感謝陳湘繁學者、林清隆學者、黃俊嘉學者、張恒嘉學者提供給我們大量的訓練資料以及蝙蝠相關的知識科普。

9. 參考文獻

[1] S. Zhang et al., "A novel bird sound recognition method based on multifeature fusion and a transformer encoder," Sensors, vol. 23, no. 19, p. 8099, 2023.

[2] Z. Li, G. Chen, and T. Zhang, "A CNN-transformer hybrid approach for crop classification using multitemporal multisensor images," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 13, pp. 847–858, 2020.

[3] D. Baroni et al., "Passive acoustic survey reveals the abundance of a low-density predator and its dependency on mature forests," Landsc. Ecol., vol. 38, pp. 1939–1954, 2023.

[4] D. Li, J. Liao, H. Jiang et al., "A classification method of marine mammal calls based on two-channel fusion network," Appl. Intell., vol. 54, pp. 3017–3039, 2024.

[5] J. C. Schäfer-Zimmermann et al., "animal2vec and MeerKAT: A self-supervised transformer for rare-event raw audio input and a large-scale reference dataset for bioacoustics," Methods Ecol. Evol., vol. 17, pp. 875–888, 2026.

[6] M. A. Tabak et al., "Automated classification of bat echolocation call recordings with artificial intelligence," Ecol. Inform., vol. 68, p. 101526, 2022.

[7] T. Mahbub et al., "Bat2Web: A Framework for Real-Time Classification of Bat Species Echolocation Signals Using Audio Sensor Data," Sensors, vol. 24, no. 9, p. 2899, 2024.

[8] E. Schwab, S. Pogrebnoj, M. Freund, F. Flossmann, S. Vogl, and K.-H. Frommolt, "Automated Bat Call Classification using Deep Convolutional Neural Networks," Bioacoustics, 2022.

[9] P. Li et al., "Learning deep models from synthetic data for extracting dolphin whistle contours," in Proc. IJCNN, 2020, pp. 1–10.

[10] J. Xie et al., "Investigation of CNN-based models for Frog Calling Activity Detection," in Proc. IEEE ICSPCC, 2020.

[11] F. Piovesan, "Exploring Transformer Model for Acoustic Scene Classification," Master's thesis, Politecnico di Torino, 2024.

[12] V. Stathopoulos, V. Zamora-Gutierrez, K. E. Jones, and M. Girolami, "Bat echolocation call identification for biodiversity monitoring: a probabilistic approach," J. R. Stat. Soc. C, vol. 67, pp. 165–183, 2018.

[13] A. Digby, M. Towsey, B. D. Bell, and P. D. Teal, "A practical comparison of manual and autonomous methods for acoustic monitoring," Methods Ecol. Evol., vol. 4, pp. 675–683, 2013.

[14] E. Kucukkulahli and A. T. Kabakus, "Towards Understanding Cat Vocalizations: A Novel Cat Sound Classification Model Based on Vision Transformers," Appl. Acoust., vol. 226, p. 110218, 2024.

[15] Q. Kong, Y. Xu, W. Wang, and M. D. Plumbley, "Sound event detection of weakly labelled data with CNN-transformer and automatic threshold optimization," IEEE/ACM Transactions on Audio, Speech, and Language Processing," vol. 28.

[16] X. Liu, "Highly efficient sound classification for marine mammals," Master's thesis, Univ. of British Columbia, 2024.

[17] B. Swaminathan, M. Jagadeesh, and S. Vairavasundaram, "Multi-label classification for acoustic bird species detection using transfer learning approach," Ecol. Inform., vol. 80, p. 102471, 2024.

[18] S. Kahl, C. M. Wood, M. Eibl, and H. Klinck, "BirdNET: A deep learning solution for avian diversity monitoring," Ecol. Inform., vol. 61, p. 101236, 2021.

[19] B. Silva, F. Mestre, S. Barreiro, P. J. Alves, and J. M. Herrera, "soundClass: An automatic sound classification tool for biodiversity monitoring using machine learning," Methods Ecol. Evol., vol. 13, pp. 2356–2362, 2022.

[20] S. M. M. Brinkløv et al., "Open-source workflow approaches to passive acoustic monitoring of bats," Methods Ecol. Evol., vol. 14, pp. 1747–1763, 2023.

[21] S. Heinicke et al., "Assessing the performance of a semi-automated acoustic monitoring system for primates," Methods Ecol. Evol., vol. 6, pp. 753–763, 2015.

[22] A. D. P. Ramirez, J. I. de la Rosa Vargas, R. R. Valdez, and A. Becerra, "A comparative between Mel Frequency Cepstral Coefficients (MFCC) and Inverse Mel Frequency Cepstral Coefficients (IMFCC) features for an Automatic Bird Species Recognition System," in Proc. IEEE LA-CCI, 2018, pp. 1–4.