

## 二、研究計畫內容：

### (一) 摘要

海洋生態系被認定是一個能夠儲存很多碳的地方，紅樹林在海岸地帶，容易經營與保育，因而成為最為重要的儲碳目標區，根據 Nelleman 等學者研究，紅樹林的碳儲存能力是所有藍碳當中最高 (Nelleman et al., 2009)。然而現有的土壤藍碳分析所需成本較高，因此，本計畫將針對在淡水挖子尾溼地的三個採樣點的土壤進行分析，將預處理完成的土壤樣本分別使用 EA 元素分析儀與 FTIR 紅外線光譜儀進行分析，然後結合 4 種 AI 模型：(1)基於規則和實例的回歸模型 (Cubist)；(2)隨機森林 (Random forest)；(3)極限學習機 (Extreme Learning Machine, ELM)；(4)偏最小平方回歸 (Partial Least Square Regression, PLS)，透過 AI 模型來學習 FTIR 數據與元素分析儀數據之間的關聯性，在這過程中嘗試不同的參數設定以找出最理想的分析結果，在模型訓練完以後，將透過三種驗證方法：(1)R 平方 (R Squared,  $R^2$ )；(2)均方根誤差 (Root Mean Square Error, RMSE)；(3)對稱平均絕對百分比誤差 (Symmetric Mean Absolute Percentage Error, sMAPE)，來判斷模型的好壞程度。接下來，根據好壞程度的依據與模型效率及效能的花費以進行比較，比較後選定一種最合理的模型。最終目的是希望能得到最理想的模型與參數設定，應用在土壤分析上透過近遠紅外光與 AI 來配合分析，能夠使土壤元素分析的時間與成本降低。

### (二) 研究動機與研究問題

#### 1. 研究動機與背景介紹

國發會於民國 111 年 3 月底發表了台灣 2050 淨零排放路徑規劃，宣示要在 2050 時達到淨零排放的目標，碳中和、碳權、自然碳匯等議題也持續升溫，如何將空氣中的二氧化碳移除並儲存在其他碳匯，已經是一個相當重要的課題，利用負碳技術成為一個很重要的碳中和方式，自然碳匯也是國發會所提出的淨零排放 12 項戰略之一(國家發展委員會, 2022)，在各種儲碳環境中「海洋生態系」被認定是一個能夠儲存很多碳的地方，因此「藍碳(Blue Carbon)」被視為一個相當重要的方向。「藍碳」這個概念首次正式由 2009 年的一次聯合國多機構(UNEP, FAO and IOC/UNESCO)聯席會議所提出，主旨為世界各國應加強研究、重視、保育與增進海岸生態系的吸碳 (Absorbing Carbon) 與封碳(Sequestering/Storing Carbon) 角色(Nelleman et al., 2009)。經過近十年的比較研究後，學者建議紅樹林的儲碳是國家層級最有效益的減碳途徑之自然解方 (Taillardat et al., 2018)。因此，為了要增加紅樹林的儲碳，勢必需要進行碳含量的分析，針對這一個部分我希望能夠引入 AI 模型，以應用於加速土壤含碳量分析時間及減少分析成本。

本計畫所使用的土壤樣本採集與聯合國糧食及農業組織所使用的方法相同 (FAO, 2020)。以下為目前的土壤採集與資料分析，對於土壤採樣的部分，土壤採樣地點位於淡水挖子尾濕地的三個採樣地點(圖 1 中，採樣點 WZW-A、WZW-B、WZW-C)，本計畫與台灣大學地質科學系羅立助理教授的團隊合作，他們已進行土壤樣本元素分析。逢甲大學團隊則進行 FTIR 光譜儀分析。三個採樣點中的一個採樣點位於沙灘區，另外兩個位於水鼻仔紅樹林區。紅樹林區的底質粒徑屬極細砂至細砂之間，適合水鼻仔生長(新北市政府農業局, 2014)。由於人工智慧模型可以協助人們處理大量、重複且繁鎖的事務，同時根據目前已有的研究，我們希望透過尋找 FTIR 分析結果與元素分析數據之間的關聯性尋找合適的 AI 模型，因此我們會研究 4 種 AI 模型，並從中挑選出最合適的模型來協助進行土壤碳含量分析，以此來加快碳分析的效率與降低成本。



圖 1. 淡水挖子尾採樣地點（台灣大學地質科學系羅立博士提供）

## 2. 研究問題

現行取得土壤碳含量前需要經過許多繁瑣的流程，同時需花費較高的費用與時間成本，如透過元素分析儀分析樣本一個樣本大約需花費 16 分鐘來完成，而使用近遠紅外線光譜分析(FTIR)時間相對快速許多。因此我們希望透過建立元素分析數據與 FTIR 光譜數據的關聯性，並且透過 AI 模型來協助建立最佳模型。我們將透過四個模型的比較與權重的調整，來找出最合理的模型與各種參數的權重設定。因此本計畫研究的重點是，如何從這四種模型當中去找到最合理的模型與參數而得到最好的結果，未來再透過 FTIR 分析並且將資訊輸入模型進行碳含量估計，以達到加快碳分析的速度與成本的降低。

### (三) 文獻回顧與探討

「碳匯」是指將二氧化碳或其他溫室氣體，如 CH<sub>4</sub>、N<sub>2</sub>O、HFCs、PFCs、SF<sub>6</sub>、NF<sub>3</sub> 等六種排放後，吸收或儲存至樹木、森林、土壤、海洋、地層、設施或場所 (全球法規資料庫, 2015)，當中「有機碳」與「總氮」對於減碳與排碳具有一定的重要性。根據參考資料 (豐年雜誌, 2022)，植物進行光合作用是土壤碳吸存的主要過程，土壤碳庫為礦物、植物的殘骸、土壤當中的分泌物與黏土礦物所聚合而成，其中有機碳會與許多微生物進行化學反應，微生物將會分解有機碳以獲取能量，並且從中攝取養分，例如氮、磷、鉀與其他微量元素。土壤碳庫中，有機碳比例明顯較多，主要是透過微生物分解/化學降解有機物而產生，是土壤碳匯研究與應用的重點。反之，剩餘的無機碳並不活躍，對於減碳、固碳上幫助有限。因此本計畫將會針對有機碳與氮分析進行探討。

參考文獻(Donato et al., 2011)中提到將二氧化碳儲存到藍碳當中，且透過維護與修復沿海生態系能夠加強碳儲存，此方法能夠減緩大氣中二氧化碳的提升與全球暖化。藍碳也就是代表儲存在海洋生態系裡面的碳 (Nelleman et al., 2009)。藍碳與熱帶森林的綠碳相比，藍碳因為環境的因素與較慢的有機物分解率，因此它擁有較高的碳儲存量 (Duarte et al., 2011)。其中紅樹林在藍碳當中被視為最重要的，紅樹林分布在熱帶與亞熱帶沿岸海域，被認為是熱帶森林中含碳量最豐富的生態系統之一 (Lugo et al., 1974)。雖然紅樹林只佔了熱帶地區面積的 0.7%，全球陸地面積的 0.1% (Giri et al., 2011)，但他的生產力能夠佔全球沿岸生態系的一半(Bouillon et al., 2008)，因為有機碳不但能儲存在樹幹中，還能夠儲存於深至 3 公尺的底土當中 (Donato et al., 2011)，由於它的高生產力、大量的碳儲存，紅樹林是一個高效儲碳的生態系。

根據參考文獻 (Li et al., 2018)，林幸助教授研究團隊量化了臺灣西海岸北、中與南部主要紅樹林生態系的碳循環，當中包括了水鼻仔與海茄苳的生物量、生產量、落葉量、碎屑輸出量以及分解速率。根據研究結果顯示，因為環境的條件優勢，台灣的紅樹林樹種密度，能夠擁有較高的儲碳能力，儲碳能力為陸地森林的 6 倍，其中有一半的碳是儲存在樹幹當中，其餘則是掉落物，像是落葉、枯枝。其中有 83% 的掉落物

在一年內會被分解，但在當中只有 6.6%會儲存在紅樹林底土內。因此，分解作用為決定紅樹林儲存量的重要因子。

根據參考文獻 (Du et al., 2009)，學者基於 PLS(偏最小平方回歸)模型分析 FTIR 光譜儀數據來推測土壤當中的氮(N)、磷(P)、鉀(K)和其他有機質含量。研究當中找出了對於 PLS 因子的定義與可能影響的因素，同時，此文獻也找出了氮與磷這兩種元素可能會是我們在分析 FTIR 結果的重要關鍵。

本研究將透過基於規則和實例的回歸模型、隨機森林、極限學習機、偏最小平方回歸等四種模型來尋找元素分析結果與 FTIR 光譜之間的關聯性。根據學習以上期刊研究的內容，能夠評估並確立何種方法來進行光譜預處理與優化決定因數，以此來增進未來研究紅樹林藍碳儲存的順暢度。

#### (四) 研究方法及步驟

##### 1. 土壤樣本處理與分析流程：

完成土壤採樣後，樣本先在實驗室進行土壤方塊烘乾與土壤切片冷乾，接著將烘乾完的土壤進行切片並且磨成粉末，最後將粉末狀之土壤樣本放入 EA 元素分析儀與 FTIR 光譜儀來進行分析。

##### 1.1 EA 元素分析儀

##### (1) 元素分析儀原理：

採用高溫動態燃燒法，當儀器運作時，燃燒管定溫為 1100~1150°C，透過重力原理垂直投入燃燒管，再以適當量的氧氣與錫之助燃下，樣品瞬間燃燒溫度將會達到 1800°C，使得樣品能夠完全燃燒，再將燃燒過後所生成的氣體於還原管當中與銅粉反應，以轉變成 $N_2$ 、 $H_2O$ 、 $CO_2$ 等氣體，再通過特殊氣體將各元素分離，最後經由熱傳導檢測器進行檢測，最後則可測得碳、氮、硫之重量百分比。

##### (2) 經由分析儀所得出之數據：

將處理好的土壤樣本放入分析儀後得出了三個採樣地點的惰性有機碳 (Recalcitrant Organic Carbon, ROC)、活性有機碳 (Labile Organic Carbon, LOC)、總有機碳 (Total Organic Carbon, TOC)、總無機碳 (Total Inorganic Carbon, TIC)、總氮 (Total Nitrogen, TN)、總碳 (Total Carbon, TC)之百分比含量，並且將其對應到每一筆數據的深度。目前的數據量為，地點 A 共有 41 筆、地點 B 共有 50 筆、地點 C 共有 32 筆，共有 123 筆數據(圖 2)。

	A	B	C	E	F	G	I	K	M
	深度	深度	深度	深度	深度	深度	深度	深度	深度
1	WZW-A	0	1	1.512528	208747.4167	95033.31	0.028101	0.018268268	1.540629
2	WZW-A	1	2	1.369324	191016.4167	83108.31	0.046805	0.025720524	1.41613
3	WZW-A	2	3	1.672323	240280.4167	95762.31	0.023273	0.027761808	1.696506
4	WZW-A	3	4	1.064286	140366.4167	65734.31	0.026201	0.013890175	1.090487
5	WZW-A	4	5	1.470845	203577.4167	91536.31	0.050288	0.035998048	1.521134
6	WZW-A	5	6	1.623674	226297.4167	101583.3	0.032668	0.010779299	1.656342
7	WZW-A	6	7	1.642143	231473.4167	102737.3	0.029726	0.049809357	1.671869
8	WZW-A	7	8	1.746545	251904.4167	107479.3	0.023264	0.090043015	1.769809
9	WZW-A	8	9	1.589213	221089.4167	101566.3	0.027503	0.031247461	1.616716
10	WZW-A	9	10	1.480027	206447.4167	94569.31	0.026604	0.026451027	1.506631
11	WZW-A	10	11	0.891448	104141.4167	65824.31	0.01693	0.000582859	0.908378
12	WZW-A	11	12	1.03544	125925.4167	84276.31	0.017832	0.030416406	1.053271
13	WZW-A	12	13	0.872988	106864.4167	63596.31	0.019599	0.00662507	0.892587
14	WZW-A	13	14	0.848692	95116.4167	69382.31	0.021034	0.007016693	0.869726
15	WZW-A	14	15	0.966579	111460.4167	74344.31	0.027503	0.010562332	0.994082
16	WZW-A	15	16	1.056657	114292.4167	90041.31	0.020488	0.000582859	1.077145
17	WZW-A	16	17	0.675035	70059.4167	53758.31	0.038871	0.021294296	0.713906
18	WZW-A	17	18	0.727564	82711.4167	56492.31	0.016887	0.007786919	0.744451
19	WZW-A	18	19	0.578114	62960.4167	39889.31	0.014325	0.002933107	0.592439
20	WZW-A	19	20	0.787391	90982.4167	58222.31	0.019499	0.009803346	0.806869
21	WZW-A	20	21	0.854628	99523.4167	63137.31	0.018993	0.000582859	0.87362
22	WZW-A	21	22	0.793238	82048.4167	60608.31	0.017121	0.019540489	0.810359
23	WZW-A	22	23	0.932843	90603.4167	89349.31	0.030698	0.00915328	0.963541
24	WZW-A	23	24	0.943728	110544.4167	79754.31	0.021686	0.010413196	0.965414
25	WZW-A	24	25	0.764409	81481.4167	60642.31	0.024805	0.005842323	0.789214
26	WZW-A	25	26	0.918478	112038.4167	67846.31	0.019693	0.000582859	0.938171
27	WZW-A	26	27	0.942304	108433.4167	69843.31	0.017606	0.02522465	0.959991
28	WZW-A	27	28	1.024905	117575.4167	78519.31	0.024167	0.009424086	1.049072
29	WZW-A	28	29	0.755435	81998.4167	64402.31	0.031337	0.007367945	0.786772
30	WZW-A	29	30	1.043067	128399.4167	84415.31	0.024016	0.015306394	1.067083
31	WZW-A	30	31	1.026009	120857.4167	78097.31	0.026495	0.000582859	1.052504
32	WZW-A	31	32	1.205743	142299.4167	97821.31	0.029847	0.019075177	1.23559
33	WZW-A	32	33	1.467809	176935.4167	120490.3	0.036065	0.013250853	1.503874
34	WZW-A	33	34	1.261843	146572.4167	104934.3	0.032977	0.009342836	1.29482
35	WZW-A	34	35	1.157922	137833.4167	105114.3	0.022105	0.013427643	1.180027

圖 2. 採樣地點 A 之元素分析數據

##### 1.2 FTIR (Fourier-Transform Infrared Spectroscopy, 傅立葉轉換紅外光譜)

##### (1) 土壤 FTIR 分析：

土壤樣本經過 FTIR 光譜儀能夠產出在不同波數(Wavenumber)上的穿

透率(Transmittance)數據 (圖 3)。

(2) 首先，我先針對 FTIR 資料進行初步分析，三個地點分別繪製同一地點不同深度的波形-穿透率之圖形以找出較特別的特徵。

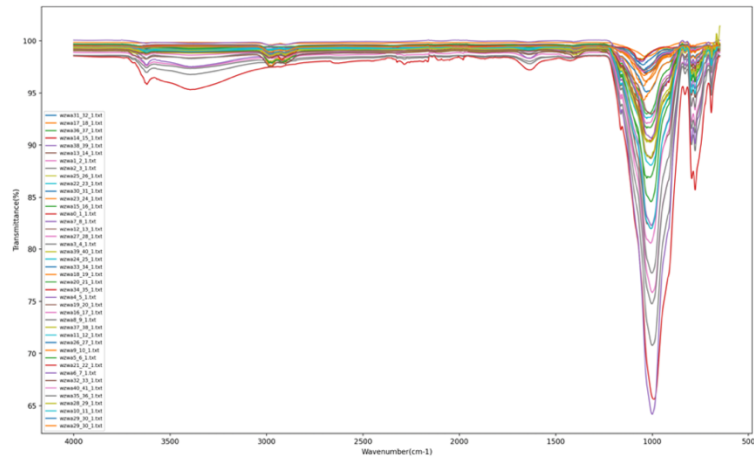


圖 3. 地點 WZW-A 之波數-穿透率圖形

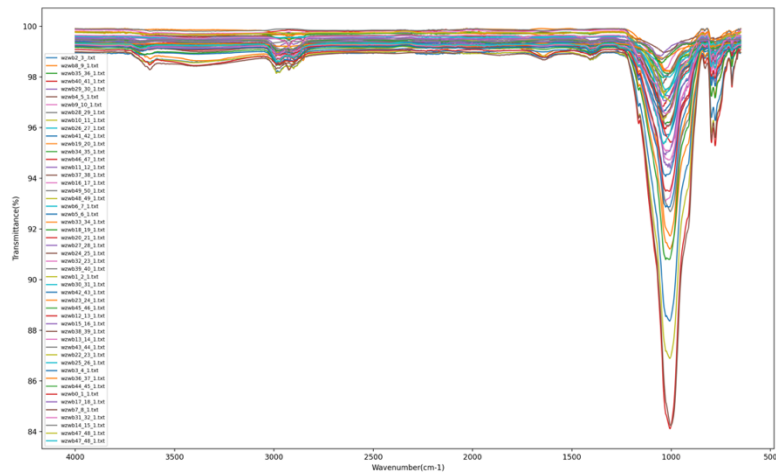


圖 4. 地點 WZW-B 之波數-穿透率圖形

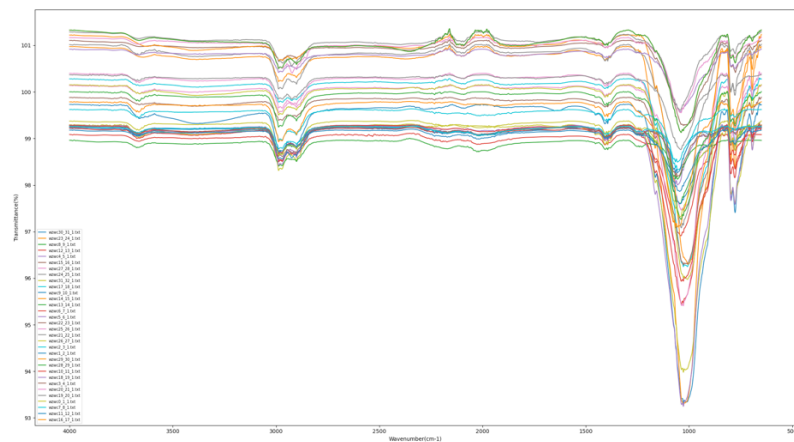


圖 5. 地點 WZW-C 之波數-穿透率圖形

由上面三個圖形可以看出 WZW-A、WZW-B 地點(紅樹林)與 WZW-C 地點(沙灘)土壤樣本的 FTIR 波譜特性有明顯的差異，暗示有機碳含量可由光譜特徵表現出來。

## 2. 土壤數據分析：

當我們數據準備完成後，將其進行數據分析，分析主要分兩個方向。第一點



，透過 FTIR 波數與穿透率之數據對應到的土壤有機成份，再將資料進行相互關聯的標記。第二點，元素分析儀所得出的數據，針對數據內容進行初步分析，再透過與 FTIR 數據所對應到的資訊，並且透過機器學習來學習兩種(元素與光譜)數據之間的關聯性。

根據參考文獻 (Coates, 2000)，紅外光光譜在波數範圍 $3300^{-1}$ ~ $2500^{-1}$ 範圍內具有較廣泛的特徵。在主要指紋庫光譜區域 ( $1500^{-1}$ ~ $400^{-1}$ )，在這個區域裡擁有較多的相似化合物。其中在波數 $1640^{-1}$ ~ $1270^{-1}$ 為氮-氧化合物 (Nitrogen-Oxygen Compounds)存在區域。

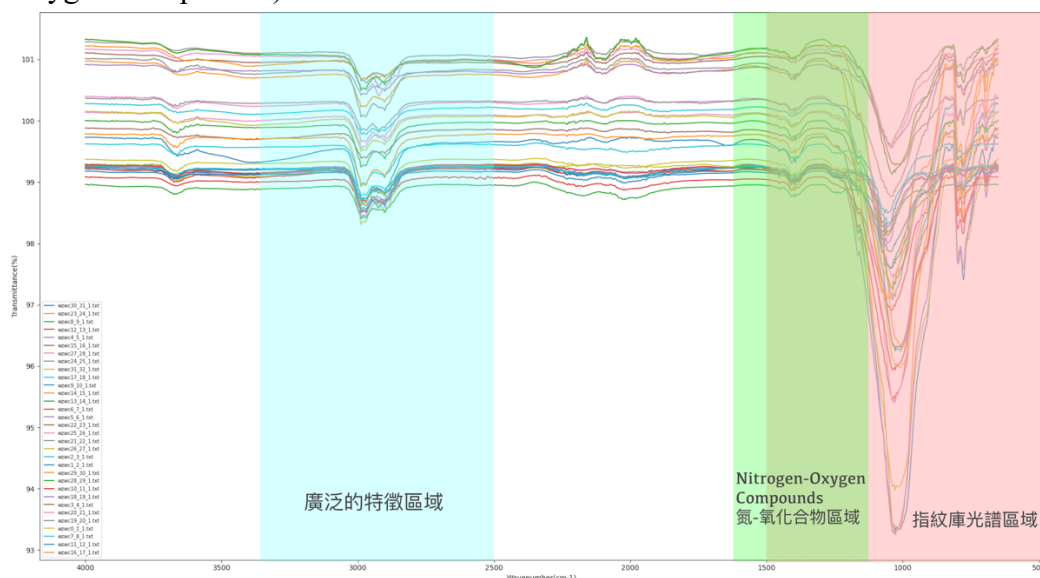


圖 6. FTIR 光譜分析圖初步分析圖。本圖標記資訊取自 Coates, 2000

### 3. 機器學習於 FTIR 數據和元素分析儀數據之分析：

人工智慧技術包含了機器學習與深度學習，在機器學習裡面又分為監督式學習 (Supervised Learning)、非監督式學習 (Unsupervised Learning) 以及強化學習 (Reinforcement Learning)。以監督式學習來講，它擁有較廣泛的應用範圍，因為它擁有明確的學習目標，也就是數據中的目標值 (Target Value) 又稱標籤 (Label)，期望能藉由預測 (Prediction) 或估計 (Estimation) 目標值。目標值有兩種類型，一種為連續數值、另一種為離散的類別，前者應用被稱為回歸，後者被稱為分類。非監督式學習，僅透過尋找特徵，也就代表它沒有學習目標，主要應用為分群 (Clustering) 與維度縮減 (Dimension Reduction)。強化學習，所做的並非為預測，而是建立一個連續決策 (Continuous Decision) 的代理人 (Agent)，因此要最佳化決策行動 (Action)，過程當中模型會根據代理人與環境 (Environment) 互動後所得到的報酬 (Reward) 與狀態 (State) 轉移的數據。模型會根據報酬的好壞自行逐步修正，以得到最佳的決策 (李家岩, 洪佑鑫, 2022)。

我的計畫將會採用監督式的模型來進行訓練，使用監督式的原因，因為我們所要尋找的目標是碳含量，根據元素分析儀與 FTIR 的數據關聯性，透過回歸的方式來尋找我們的目標值。

#### 3.1 AI 模型

##### (1) 基於規則和實例的回歸模型 (Cubist)

Cubist 是由 John Ross Quinlan (RuleQuest Research) 所開發的一種回歸演算法，主要適用於生成基於規則的預測模型。它使用了集成學習的想法，將多個回歸模型組合在一起，模型會透過對訓練集當中的資料尋找他們之間的關聯性，以建立一組回歸模型。如圖 9 所示，此圖裡面所要預測的目標為 outcome，Cubist 會去學習數據當中的其他項目與目標的關聯性並且生

成出方程式。以圖 7 規則 1 為例，他的規則是假如 petal\_length 這個項目的數值小於等於 4.7，就將各項目的數值帶入模型生成出的 outcome 預測方程式 (Aselin, 2021)。

- 優點：Cubist 可以透過增加更多的模型來增加預測的準確度，這些模型可以針對先前的錯誤做修正。
- 缺點：可能會有過度擬和的情形發生，對於參數選擇需要適當地去調整。訓練所需時間較長。

選用 Cubist 的主要原因是因為此模型是去尋找數據之間的關聯性，與本計畫類型相同，透過此模型來去尋找 FTIR 與元素分析儀數據之間的關聯性。

圖 7. Cubist 模型規則示意。本圖取自 Aselin, 2021

## (2) 隨機森林 (Random Forest)

- 模型介紹：

隨機森林 (Brownlee, 2016) 是一種基於決策樹分類器的集成式學習演算法，由 L. Breiman 在 2001 年提出。隨機森林演算法中的分類器為 CART (Classification And Regression Tree)，使用 Bagging (又稱 Bootstrap Aggregating) 演算法進行組合學習，並在 CART 樹進行生長時隨機選取變數以進行分類。CART 演算法概念是使用遞迴二元分割 (Recursive Binary Splitting) 來歸納與分析大量複雜變數的資料集。

- 樹生成方法：

第一步，從訓練集當中抽取  $n$  筆資料(這  $n$  筆資料能夠重複抽取)。以我們目前數據為例，假設有 123 筆，從中抽取 50 筆數據，這 50 筆資料可能會有重複的數據。第二步，從這些資料當中挑選  $i$  個特徵當作決策因子，代表每一棵樹都只能看到部分的特徵。最後，重複以上步驟  $k$  次並產生  $k$  棵樹。當建立多棵決策樹後，將它們的預測結果結合起來。若為分類問題，則採用多數決的方法來整合預測結果，若為回歸問題，則採用平均法來整合預測結果。

- 優點：

- 因為隨機森林是基於 CART 演算法，故可以處理類別資料與連續資料。
- 透過集合多組決策樹的預測結果來提升準確度。
- 較不容易過度擬和。
- 能夠接受高維度的特徵資料

- 缺點：

- 運算時需要大量的資源，以儲存每棵樹的資訊。
- 因隨機森林是將多棵決策樹組和起來，因此無法對單一樹作解釋。

## (3) 極限學習機 (Extreme Learning Machine, ELM)

- 模型介紹：

極限學習機 (Huang et al., 2004) 是一種單隱藏層前潰神經網路

(SLFN)學習演算法，這種演算法只需要設置網路的隱藏層節點數量與激活函數。執行過程中不需要調整網路的輸入權重，在輸入層到隱藏層的權重是透過隨機產生，然後再使用最小平方方法來確定輸出層的權重。ELM結構主要由兩層網路所組成，第一層先隨機初始化然後固定，第二層則是進行訓練。

- 訓練流程：

第一步：先輸入訓練集  $\mathbf{X} = \{(x_i, t_i) | x_i \in R^n, t_i \in R^m, i = 1, \dots, N\}$ ，激活函數  $g(x)$  與隱藏層數量  $\tilde{N}$

第二步：模型隨機生成隱藏層權重  $w_i$  與偏誤  $(b_i, i = 1, \dots, \tilde{N})$

第三步：計算隱藏層的輸出矩陣  $H$

第四步：使用最小平方方法來確定輸出的權重  $\beta$

- 優點：ELM 演算法可以在短時間內訓練完成且精準度高
- 缺點：解釋性較低，可能會出現過度擬和的情況

#### (4) 偏最小平方回歸(Partial Least Square Regression, PLS)

偏最小平方回歸是一種基於協方差的統計學方法。針對 PLS 在模型建立上，與多元線性回歸、主成分分析及其他重要的分析技術有關。PLS 是一種用來建構預測模型的統計方法 (Tobias, 1996)。

### 3.2 AI 模型之準確度驗證

針對本計畫所使用到的模型，為了驗證模型的好壞程度與回歸方程式的合理性，我將採用以下 3 種方法 (李家岩, 洪佑鑫, 2022)：R 平方(R Squared,  $R^2$ )；均方根誤差 (Root Mean Square Error, RMSE)；對稱平均絕對百分比誤差 (Symmetric Mean Absolute Percentage Error, sMAPE)。以下式子當中，定義實際值為  $y$ 、預測值為  $\hat{y}$ 、平均數為  $\bar{y}$ 、樣本數為  $n$ 。

#### (1) R 平方 (R Squared)：

R 平方又稱判定係數(Coefficient Of Determination)，是用來評估回歸模型的表現，代表 X 解釋 Y 的能力，X 能夠解釋 Y 變動的比例。針對本計畫，將透過 R 平方來檢驗模型所建立出的線性回歸式的合理程度。選用 R 平方的原因，因為本計畫所使用的模型均為回歸模型，R 平方就是一種衡量回歸模型表現的指標。

- R 平方公式：透過 1 減去總平方和分之殘差平方和。式(1)中的 SSR(Sum Of Square Regression)為可解釋變異(Explained Variation)是預測值  $\hat{y}$  與平均值  $\bar{y}$  相差的平方和。SST(Sum Of Square Total)為整體變異，是實際值  $y$  與平均值  $\bar{y}$  相差的平方和。SSE(Sum Of Square Error)為不可解釋變異(Unexplained Variation)，將實際值  $y$  與預測值  $\hat{y}$  相差的平方和，如公式(1)所示。

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (1)$$

- 殘差平方和(Sum Of Square Error, SSE)：透過真實數值  $y_i$  - 預測值  $\hat{y}_i$  即為殘差，為了計算方便，累加同時取平方。因此殘差平方和越大，也就代表模型解釋力越低，如公式(2)所示。

$$SSE = \sum_{i=1}^N (\hat{y}_i - y_i)^2 \quad (2)$$

- 總平方和(Sum Of Square Total, SST)：實際值  $y$  與平均值  $\bar{y}$  相差的平方和，如公式(3)所示。

$$SST = \sum_{i=1}^N (y_i - \bar{y})^2 = SSR + SSE \quad (3)$$

$$= \sum_{i=1}^N (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^N (\hat{y}_i - y_i)^2$$

(2) 均方根誤差 (Root Mean Square Error, RMSE)：

RMSE 是一種用於評估回歸模型的好壞損失函數，RMSE 主要就是拿 MSE 去取根號，最後所得結果解釋起來會較直觀。MSE 也就是拿預測的值與真實數據相減，所得的相減值平方過後並加總，最後再除以總數量平均，因此當 MSE 值越小，說明預測模型具有較好的精準度，如公式(4)所示。

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (4)$$

(3) 對稱平均絕對百分比誤差 (Symmetric Mean Absolute Percentage Error, sMAPE)：

MAPE 是一種用於評估回歸模型的損失函數，透過計算預測值與真實數值之間的百分比誤差再取平均，將誤差轉換為百分比，主要對於解釋起來較直觀，也不需要去注意單位的問題，但 MAPE 存在三個缺點，第一點，分母為 0 時無法相除，第二點，若預測值誤差很大，指標可能會超過 100%，第三點，實際值與預測值互換時，指標會產生不對稱的結果，如公式(5)所示。

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (5)$$

因此我將選用 sMAPE，它是將 MAPE 的分母的分母進行修正為了解決不對稱的問題。整體來說，MAPE 對於同樣的實際值平衡較好，sMAPE 對於不同樣的實際值平衡較好，因此 sMAPE 對整體的誤差評估較為均衡，如公式(6)所示。

$$sMAPE = \frac{100\%}{n} \sum_{i=1}^n \frac{|\hat{y}_i - y_i|}{\frac{(|\hat{y}_i| + |y_i|)}{2}} \quad (6)$$

## (五) 預期結果

### 1. AI 模型產出：

本計畫將測試與分析適合用於 FTIR 結果與元素分析結果之對應的 AI 模型，由於 AI 模型除了較合理的演算法，更重要的為適切的參數設定，因此，找出最合理的演算法與參數設定。

### 2. 土壤元素特徵選取：

由於將土壤樣本經化學元素分析後的結果包含許多的參數，然而在訓練 AI 模型時，不一定會使用到所有的參數，因此，本計畫預期會找出較合適的特徵以利於 AI 模型的建立。

### 3. FTIR 特徵選取：



FTIR 主要是經由近遠紅外線打向受測物後觀測經由光譜的情況所得到的數據，FTIR 所得出的數據通常是以波形呈現，本計畫預期找出 FTIR 波形與元素分析結果的關聯性，以協助後續的 AI 模型建立。

#### 4. 程式能力提升：

由於現有的 AI 模型在使用上可能出現沒辦法完全涵蓋的情況，因此在本計畫執行時，將會有效訓練我們在模型的調整與設計上，並完成計畫目標所要尋找較合理的 AI 模型與建立。

#### 5. 計畫成果分享與發表：

本研究計畫除了瞭解 FTIR 結果與元素分析結果之間的關係、AI 模型建立、參數調整...等項目以外，我們將會進一步的把研究過程及成績彙整成學術論文並投稿到國際學術研討會進行發表，如獲接受發表，相信可以提升我國在應用 AI 模型於土壤分析相關研究與應用之國際能見度。

### (六) 需要指導教授指導內容

#### 1. 土壤分析/藍碳分析：

由於土壤與藍碳都並非是我有接觸過的領域，對於一些詳細內容的理解與土壤分析數據裡面數值所代表的意義，將會請教魏國彥教授(前國立台灣大學地質科學系教授)。

#### 2. 數據預處理與光譜數據特徵觀察：

針對光譜數據，如何去進行標準化以及去異質化，對於如何去驗證該種方法對於數據所產生的好、壞影響，該透過哪種方法去呈現並且進行調整，以及在光譜數據當中應該如何去評估當中的特徵，這部分將會請教梁辰睿教授(逢甲大學材料科學與工程學系助理教授)。

#### 3. AI 模型選擇與概念理解：

針對光譜數據，透過什麼方法能夠抓取特徵，對於數據之間的關聯性可以透過什麼樣的方式來標記，這方面將會請教指導教授，針對程式撰寫標記方法與如何分析關聯性。

#### 4. 光譜數據與元素分析數據特徵抓取與關聯性分析：

對於不同的數據類型與型態都會對應到不一樣的 AI 模型，AI 模型的選擇上，以我的研究計畫的資料類型來看，對於要使用何種模型較為合理並且對於 AI 模型的運作方法，需要指導教授進行相關概念的指導。

#### 5. AI 模型驗證與調整的建議：

針對本計畫所選定的 AI 模型表現程度，需要透過驗證的方法來決定，在驗證模型好、壞以後，對於模型互相的比較也包含了的運作效率與效能消耗程度的比較。對於模型權重的設定、參數的調整、模型是否過度擬和等等，AI 模型的驗證與如何調整 AI 模型需要指導教授的指導。

### (七) 參考文獻

土壤碳庫與淨零排放 降低氣候變遷衝擊最有潛力的方式. (2022). 豐年雜誌.

<https://www.agriharvest.tw/archives/90617>

李家岩, & 洪佑鑫 . (2022). 製造數據科學：邁向智慧製造與數位決策. 前程文化.

[https://www.books.com.tw/products/0010922085?loc=P\\_0020\\_004](https://www.books.com.tw/products/0010922085?loc=P_0020_004)

新北市挖子尾自然保留區生態資源監測工作成果報告書. (2014). 新北市政府農業局

<https://wetland-tw.tcd.gov.tw/upload/file/20190604135503952.pdf>

溫室氣體減量及管理法. (2015, July 1). 全國法規資料庫.

<https://law.moj.gov.tw/LawClass/LawAll.aspx?PCode=O0020098>

臺灣 2050 淨零排放路徑及策略總說明. (2022). 國家發展委員會.

- <https://ws.ndc.gov.tw/Download.ashx?u=LzAwMS9hZG1pbmlzdHJhdG9yLzEwL3JlbGZpbGUVvMC8xNTA0MC8yZTZhZTA0Mi0wYjUyLTQ4OTAtOGY5NC1hYjk5MzgZlZlucGRm&n=6le654GjMjA1MOa3qOmbtuaOkuaUvui3r%2bW%2bkeWPiuetlueVpee4veiqquaYji5wZGY%3d&icon=.pdf>
- Aselin, P. 2021. Cubist. [\(2021\)](https://github.com/pjaselin/Cubist).  
An Overview of Cubist. (n.d.). RuleQuest Research.  
<https://www.rulequest.com/cubist-win.html>
- Brownlee, J. (2016). Master Machine Learning Algorithms: Discover How They Work and Implement Them From Scratch. Machine Learning Mastery.  
[https://books.google.com.tw/books?id=n--oDwAAQBAJ&hl=zh-TW&source=gbs\\_navlinks\\_s](https://books.google.com.tw/books?id=n--oDwAAQBAJ&hl=zh-TW&source=gbs_navlinks_s)
- Bouillon S, Borges AV, Castañeda- Moya E et al. (2008) Mangrove production and carbon sinks: a revision of global budget estimates. *Global Biogeochemical Cycles*, 22, 1–12.  
<https://doi.org/10.1029/2007GB003052>
- Coates, J. (2000). Interpretation of infrared spectra, a practical approach.  
<https://doi.org/10.1002/9780470027318.a5606>
- Duarte C, Agustí S (2011) Rapid carbon cycling in the oligotrophic ocean. *Biogeosciences Discussions*, 8, 11661–11687.  
[https://www.researchgate.net/publication/228521803\\_Rapid\\_carbon\\_cycling\\_in\\_the\\_oligotrophic\\_ocean](https://www.researchgate.net/publication/228521803_Rapid_carbon_cycling_in_the_oligotrophic_ocean)
- Du, C., Zhou, J., Wang, H., Chen, X., Zhang, J., & Zhu, A. (2009). Determination of Soil Properties Using Fourier Transform Mid-Infrared Photoacoustic Spectroscopy. *Vibrational Spectroscopy*, 49(1), 32–37.  
[https://www.sciencedirect.com/science/article/abs/pii/S0924203108000805?casa\\_token=ADrFU01xYY0AAAAA:wxdySxSA5p8EajEthY6e5zAG6qwTb34h2Lx-6dNhDooFzGbDOvHJFZsoHpXWh7jnmb7jR711V21B](https://www.sciencedirect.com/science/article/abs/pii/S0924203108000805?casa_token=ADrFU01xYY0AAAAA:wxdySxSA5p8EajEthY6e5zAG6qwTb34h2Lx-6dNhDooFzGbDOvHJFZsoHpXWh7jnmb7jR711V21B)
- Donato, D., Kauffman, J., Murdiyarso, D. et al. Mangroves among the most carbon-rich forests in the tropics. *Nature Geosci* 4, 293–297 (2011).<https://doi.org/10.1038/ngeo1123>
- Fao/gsp. (2020). Soil Testing Methods Manual. FAO.  
<https://www.fao.org/documents/card/en/c/CA2796EN>
- Giri C, Ochieng E, Tieszen LL et al. (2011) Status and distribution of mangrove forests of the world using earth observation satellite data. *Global Ecology and Biogeography*, 20, 154– 159.  
<https://doi.org/10.1111/j.1466-8238.2010.00584.x>
- Huang, G., Zhu, Q., & Siew, C. Kheong. (2004). Extreme Learning Machine: A New Learning Scheme of Feedforward Neural Networks. *Proceedings of International Joint Conference on Neural Networks (IJCNN2004)*, 2.  
[https://www.researchgate.net/publication/4116697\\_Extreme\\_learning\\_machine\\_A\\_new\\_learning\\_scheme\\_of\\_feedforward\\_neural\\_networks](https://www.researchgate.net/publication/4116697_Extreme_learning_machine_A_new_learning_scheme_of_feedforward_neural_networks)
- Li, S.-B., Chen, P.-H., Huang, J.-S., Hsueh, M.-L., Hsieh, L.-Y., Lee, C.-L., & Lin, H.-J. (2018). Factors regulating carbon sinks in mangrove ecosystems. *Global Change Biology*, 24(9), 4195–4210. doi:10.1111/gcb.14322 <https://doi.org/10.1111/gcb.14322>
- Lugo AE, Snedaker SC (1974) The ecology of mangroves. *Annual Review of Ecology and Systematics*, 5, 39–64. <https://doi.org/10.1146/annurev.es.05.110174.000351>
- Nelleman, C.; Corcoran, E.; Duarte, C.M.; Valdés, L.; DeYoung, C.; Fonseca, L.; Grimsditch, G. (Eds.) *Blue Carbon: A Rapid Response Assessment*; United Nations Environmental Programme and GRID-Arendal: Arendal, Norway, 2009.  
<https://digitallibrary.un.org/record/673428>
- Taillardat, P., Friess, D. A., & Lupascu, M. (2018). Mangrove blue carbon strategies for climate change mitigation are most effective at the national scale. *Biology Letters*, 14(10), 20180251.  
<https://doi.org/10.1098/rsbl.2018.0251>
- Tobias, R.D. (1996). An Introduction to Partial Least Squares Regression.  
<https://www.semanticscholar.org/paper/An-Introduction-to-Partial-Least-Squares-Regression-Tobias/be750f3f9a875ed713999686b7b845a594c4de36#citing-papers>