

計畫編號：SWCB-111-058

**應用長短期記憶網路結合環境因子於坡地
崩塌風險評估**

**Application of long short-term memory
network combined with environmental
factors in slope collapse risk assessment
(成果報告)**

執行單位：國立陽明交通大學

執行期間：111 年 1 月 18 日至 111 年 12 月 31 日

計畫主持人：林子剛 教授

行政院農業委員會水土保持局 編印

中華民國 111 年 12 月

(本報告書內容及建議純屬執行單位意見，僅供本局施政參考)

應用長短期記憶網路結合環境因子於坡地崩塌風險評估

摘要

臺灣位處亞熱帶地區與板塊交界帶，因此有長時間降雨的梅雨季以及颱風來襲時的強降雨，並經常有地震事件的發生，兩者經常相互的影響，造成強降雨連帶產生崩塌地質災害，或是地文特性使得地震事件發生時容易發生崩塌事件。因此本計畫範圍選定台中市后里區、花蓮縣秀林鄉、新北市石碇區、南投縣信義鄉及嘉義縣竹崎鄉崩塌事件較為豐富的區域，以環境因子分析結合人工智慧相關技術，在強降雨或是強震發生後，以人工智慧模型來預測各區域潛在的崩塌發生機率，以即時做出災害應變措施以及預先的決策準備。最初階段預測模型已完成，目前模型用於崩塌預測之因子包含由長短期記憶模型(Long Short-Term Memory, LSTM)解析出之時間序列因子：降雨歷時因子值、地震強度因子值，以及環境因子：坡度、坡向、高程以及植被覆蓋狀況(NDVI)。此崩塌預測模型以 LSTM 預測之降雨及地震因子，結合環境因子對 20m*20m 之網格崩塌與否，預測準確率達到 73.8%。崩塌預測模型能夠依據過去降雨和地震資訊，以及該地地質特性、地表情況所構成之環境因子，以網格為單位進行崩塌機率預測，

提供可能的受災範圍與規模，達到即時崩塌預警與救災的目的。

關鍵詞：長短期記憶網路、類神經網路、坡地崩塌預測

Application of long short-term memory network combined with environmental factors in slope collapse risk assessment

Abstract

Taiwan is in a subtropical region, where the continental cold air mass and the Pacific warm air mass interact, causing the plum rainy season with prolonged rainfall and heavy rainfall due to typhoons. In addition to precipitation, Taiwan is located at convergent plate boundaries, and frequent earthquakes result in stratigraphic belts. These areas are prone to geological disasters such as slope collapse due to heavy rainfall and earthquake, resulting in massive loss of life and property., leading to frequent earthquakes. Because of the terrestrial characteristics, slope collapse events frequently happen when typhoons strike or rainfall intensity increases. This project selects areas with collapse events in the Houli District of Taichung City, Xiulin Township of Hualien County, Shiding District of New Taipei City, Xinyi Township of Nantou County, and Zhuqi Township of Chiayi County. Analyzing environmental factors in the historical collapse events is combined with artificial intelligence-related technologies. After heavy rainfall or strong earthquakes occur, artificial intelligence models are used to predict the potential collapse probability of each area to facilitate immediate disaster response measures and advance decision-making preparations. The factors of the proposed model include time series factors derived from the rainfall and earthquake

time history using the long-short-term memory (LSTM) model and environmental factors, including slope, aspect, elevation, and vegetation coverage based on the normalized difference vegetation index (NDVI). The proposed AI-based approach consists of two models. The first model is the LSTM model, which analyzes rainfall and earthquake time history and predicts rainfall and earthquake factors indicating the slope collapse risk. The second model is the ANN model, which incorporates rainfall and earthquake factors with environmental factors to predict the slope collapse probability of a 20m*20m grid. The proposed model conducts grid-based slope collapse probability prediction using rainfall and earthquake time history with environmental factors and reaches 73.8% accuracy for slope collapse prediction. This model can be used in regional slope collapse risk assessment and provide real-time warning for risk reduction purposes.

Keywords: Artificial neural network, Long short-term memory, slope collapse risk assessment

目次

摘要.....	I
Abstract.....	III
目次.....	V
表次.....	VII
圖次.....	VIII
第一章 緒論	1-1
第一節 計畫欲解決的問題	1-1
第二節 計畫範圍	1-3
第三節 執行流程與計畫目標	1-3
第二章 文獻回顧	2-1
第一節 類神經網路 (Artificial Neural Network).....	2-1
第二節 長短期記憶模型 (Long Short-Term Memory, LSTM)	2-3
第三章 模型訓練資料庫建立	3-1
第一節 選定研究範圍	3-1
第二節 基本資料庫說明	3-1
第三節 資料量化與因子解算	3-3
第四節 DEM 模型 20m x 20m 網格	3-6
第五節 網格資料庫建置	3-7
第六節 崩塌坡地資料庫建立成果	3-11
第四章 模型測試與訓練	4-1
第一節 模型建置流程	4-1
第二節 LSTM 網路建置.....	4-1
第三節 網路訓練和驗證	4-3
第四節 網路測試以及優化	4-5

第五節 LSTM 模型建置與訓練結果.....	4-7
第六節 ANN 模型建置與訓練結果.....	4-17
第七節 小結	4-22
第五章 結論與未來展望	5-1
第一節 結論	5-1
第二節 未來展望	5-2
參考文獻.....	參-1
附錄一、期初審查會議紀錄暨回覆辦理情形	附錄-1
附錄二、期中審查會議紀錄暨回覆辦理情形	附錄-6
附錄三、期末審查會議紀錄暨回覆辦理情形	附錄-10

表次

表 3-1 網格降雨時間序列資料.....	3-13
表 3-2 網格地震震度時間序列資料.....	3-15
表 3-3 數值化呈現的網格坡向.....	3-17
表 3-4 數值化呈現的網格坡度.....	3-18
表 3-5 數值化呈現的網格高程.....	3-19
表 3-6 網格 NDVI 值.....	3-21
表 4-1 降雨與地震震度預測因子值.....	4-18
表 4-2 ANN 訓練資料.....	4-19
表 4-3 網格預測崩塌機率.....	4-20
表 4-4 混淆矩陣數據.....	4-21

圖次

圖 1-1 莫拉克颱風當月雨量資料.....	1-1
圖 1-2 研究區分布範圍	1-3
圖 1-3 計畫執行流程圖	1-4
圖 2-1 類神經網路架構示意圖	2-1
圖 2-2 FNN 模型架構圖	2-2
圖 2-3 遞迴式神經網路架構示意圖	2-3
圖 2-4 LSTM 網路架構示意圖	2-4
圖 3-1 QGIS 操作介面示意圖.....	3-2
圖 3-2 72 小時之時雨量資料示意圖.....	3-4
圖 3-3 降雨因子值設定方法.....	3-4
圖 3-4 地震歷史紀錄資料示意圖	3-5
圖 3-5 震強度因子值設定方法.....	3-5
圖 3-6 網格建置示意圖	3-7
圖 3-7 20M X 20M 網格示意圖	3-7
圖 3-8 坡向資料	3-8
圖 3-9 坡度資料	3-9
圖 3-10 高程資料	3-10
圖 3-11 SENTINEL-2 衛星之 NDVI 圖	3-10
圖 3-12 花蓮縣帕托魯山附近的崩塌地區示意圖	3-11
圖 3-13 花蓮縣帕托魯山附近的崩塌範圍設定示意圖	3-12
圖 3-14 花蓮 6/2~6/4 的第 1 格網格(未崩塌)降雨時間序列.....	3-13
圖 3-15 花蓮 6/2~6/4 的第 20 格網格(崩塌)降雨時間序列	3-14
圖 3-16 地震測報中心的震度資料.....	3-15
圖 3-17 花蓮 6/2~6/4 的第 1 格網格(未崩塌)地震震度時間序列 ..	3-16

圖 3-18 花蓮 6/2~6/4 的第 20 格網格(崩塌)地震震度時間序列....	3-16
圖 3-19 臺灣坡向分布圖	3-17
圖 3-20 花蓮地區網格坡向圖	3-17
圖 3-21 花蓮地區網格坡度圖	3-18
圖 3-22 HDR 檔案中的網格平均高程數據.....	3-19
圖 3-23 臺灣高程陰影圖	3-20
圖 3-24 花蓮 NDVI 圖	3-21
圖 4-1 模型流程圖	4-1
圖 4-2 LSTM 模型架構概述示意圖	4-3
圖 4-3 RELU 激發函數示意圖	4-3
圖 4-4 SIGMOID 激發函數示意圖.....	4-3
圖 4-5 訓練曲線圖收斂示意圖	4-4
圖 4-6 驗證曲線圖收斂示意圖	4-5
圖 4-7 崩塌機率預測結果可視化.....	4-6
圖 4-8 降雨因子值設定方法	4-9
圖 4-9 地震震度因子值設定方法	4-9
圖 4-10 第一次降雨訓練圖	4-10
圖 4-11 第一次地震震度訓練圖	4-10
圖 4-12 第二次降雨訓練圖	4-11
圖 4-13 第二次地震震度訓練圖	4-11
圖 4-14 第三次降雨訓練圖	4-12
圖 4-15 第三次地震震度訓練圖	4-12
圖 4-16 第四次降雨訓練圖	4-13
圖 4-17 第四次地震震度訓練圖	4-14
圖 4-18 第五次降雨之訓練圖	4-15

圖 4-19 第五次地震震度訓練圖	4-15
圖 4-20 降雨 LSTM 訓練集	4-16
圖 4-21 降雨 LSTM 測試集	4-16
圖 4-22 地震震度 LSTM 訓練集	4-17
圖 4-23 地震震度 LSTM 測試集	4-17
圖 4-24 混淆矩陣圖	4-20
圖 4-25 ROC 曲線圖	4-21
圖 4-26 崩塌預測圖	4-22

第一章 緒論

臺灣位處於板塊交界帶，地震的頻發使得該區域之地層敏感度高，當強降雨事件發生時容易造成崩塌地質災害的發生，2009 年莫拉克颱風來襲，在短短幾天內便下了臺灣一整年的降雨量(如圖 1-1)，因為持續性降雨使得溪水暴漲、多處崩塌以及大規模山崩等連續發生之地質災害造成多達 474 人遭到活埋。因此本計畫使用降雨歷時資料以及地震強度歷史資料做為時間序列因子，坡度、高程、坡向、植被覆蓋狀況等做為環境因子，藉由量化分析此兩種因子做為人工智慧模型的輸入資料，能以分割單位為目標進行崩塌積率的預測，提供災害應對的參考依據，並即時的做出更有效的災害應對措施。

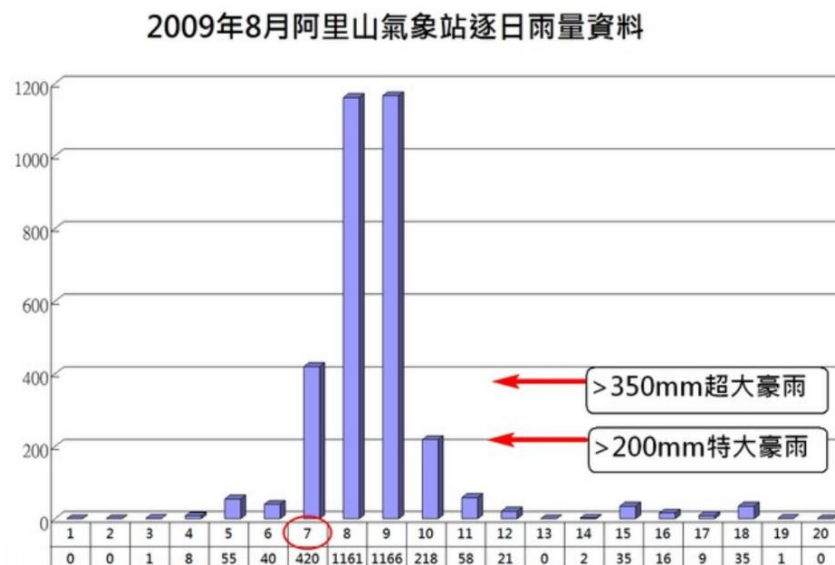


圖 1-1 莫拉克颱風當月雨量資料

資料來源：小林村滅村的原因:我的觀點

第一節 計畫欲解決的問題

坡地崩塌的發生特性受現地岩石特性、地質構造、土地利用情形、

水文情況等多樣的環境因子所影響，且強降雨、地震等突發因素更是崩塌發生的誘發因子。一旦山區發生崩塌，造成建築倒塌或聯外道路受阻，將使救災及援助變得極為困難。因此，強降雨情形出現之即時災害發生預測以及區域地震歷史對於該地的潛在影響，都是預防崩塌發生所要探討的問題。本計畫將結合降雨歷時資料以及地震歷史資料，結合當地地文特性，進行崩塌潛勢機率預測，找出可能的受災範圍與規模，提供即時的警報與第一時間的救援，保證人民的生命財產安全與降低崩塌所帶來的威脅。

為預防以及減少崩塌帶來的損害，本計畫擬以人工智慧協助預測崩塌發生的潛在機率，實現即時的災害評估。該人工智慧模型能對每單位網格進行崩塌機率預測，模型輸入為時間序列因子和環境因子，其中時間序列因子需由一長短期記憶模型(Long Short-Term Memory, LSTM)所判定。時間序列資料有一定的潛在特性是人為判讀難以發覺，每一筆時間序列資料之型態會受其特性所影響，例如：降雨延時、降雨強度等等，而 LSTM 根據降雨資料以及歷史地震分布時間序列資料之特性來找出該時間序列之潛在特性並判定時間序列因子。隨後結合該因子以及其他環境因子做為計算崩塌機率的輸入資料，以該輸入資料做為前饋式神經網路的訓練資料建置一崩塌機率預測模型，其中環境因子包含地質特性產生之地文因子。

以人工智慧技術建立一崩塌機率預測模型，本團隊進行一系列資料收集以及模型建置與測試流程，透過資料的量化與分析結合人工智慧相關技術建立一有效之預測模型，提供各區域崩塌發生機率預測之有效依據。

第二節 計畫範圍

本計畫係針對多個區域，包含台中市后里區、屏東縣車城鄉、花蓮縣秀林鄉、新北市石碇區、南投縣信義鄉及嘉義縣竹崎鄉(如圖 1-2)，進行資料收集以及量化分析，這些地區為崩塌事件數量較多之區域，因此被選定為這次研究範圍。

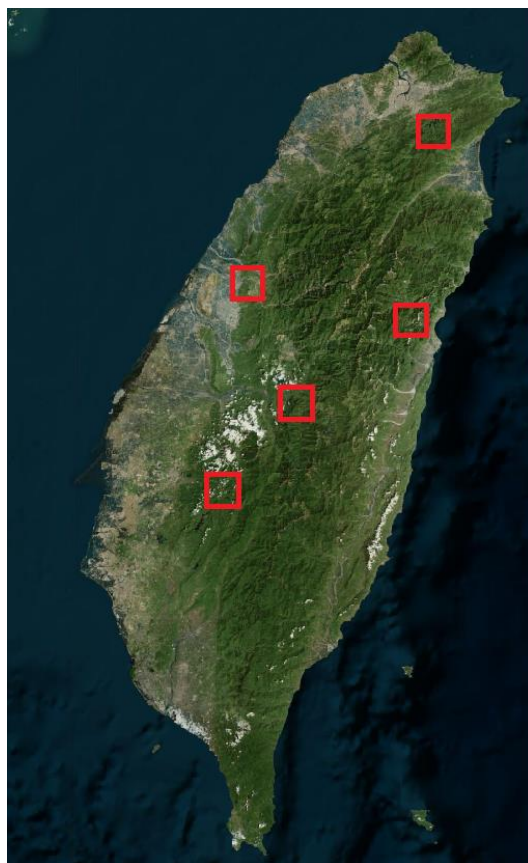


圖 1-2 研究區分布範圍

資料來源：本計畫繪製

第三節 執行流程與計畫目標

本計畫主要配合水土保持局 110 年度創新研究計畫之「基礎試驗研究」領域，期以人工智慧結合以及環境因子，探討不同觸發條件對不同坡地條件上的影響，並預測強降雨發生或地震發生後，崩塌潛在

的發生機率，以利即時的做出災害應變措施以及提供未來調整坡地災害警戒發布值之依據。

透過收集坡地之地文資料以及時間序列資料，建立基本資料庫，並利用 LSTM 模型對時序資料解算出時間序列因子。結合地文資料整理而成之環境因子以及解算出之降雨時序因子、地震歷史時序因子，利用前饋式類神經網路來計算每個大小為 20m*20m 之網格崩塌機率，達成坡地崩塌風險評估，以利協助強降雨或地震發生時崩塌發生之預警，並提供更即時的救災規劃資訊，降低坡地崩塌的負面影響以及保障人民的財產生命安全，執行流程如圖 1-3 所示。

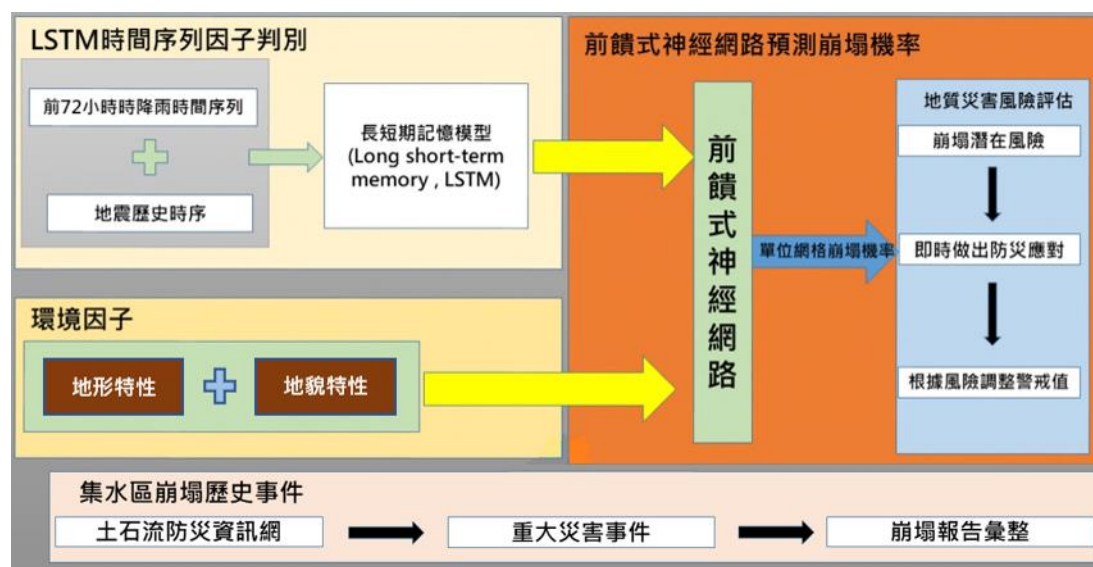


圖 1-3 計畫執行流程圖

資料來源：本計畫繪製

預期達成之計畫目標如下。

- 一、資料蒐集整理：從各機關資料開放平台(水土保持局資料開放平台、國土測繪中心、觀測資料查詢系統(CWB Observation Data Inquire System, CODiS)、中央氣象局地震測報中心、內政部地籍

圖資網路便民服務系統等)或文獻中蒐集研究範圍內各類地文資料以及時間序列資料。

二、建立基本資料庫：運用QGIS等地理資訊軟體，將各類圖資(高程圖、坡向圖等)進行數值化，結合時間序列資料(降雨歷時、地震歷史紀錄)，整理並記錄上述地文資料和時序資料，並建立長短期記憶模型。以時序資料訓練模型，預測出時間序列因子。

三、風險評估成果檢視：建立前饋式類神經網路模型，使用資料庫中崩塌發生之歷史事件為訓練資料，訓練一類神經網路模型之權重，使其能準確結合各項因子來預測崩塌機率。使用的因子包括地文因子以及由長短期記憶模型所得之時間序列因子，進行坡地發生崩塌之機率預測並顯示於網格資料中。

四、風險評估成果檢視：評估模型準確度，並說明該模型可能運用範圍以及未來展望，以期該研究成果成為災害預警及即時救災之依據。

第二章 文獻回顧

第一節 類神經網路 (Artificial Neural Network)

類神經網路是研究者由人類中樞神經系統得到的啟發，已被稱做神經元(neurons)的人工節點組成一個類似生物神經網路的網狀結構。神經網路架構通常被分成三層(圖 2-1)，分別是輸入層、隱藏層、輸出層，輸入層是由多個能接受輸入資料的神經元組成，隱藏層為神經網路學習的核心，介於輸入層和輸出層之間，使用非線性的轉換函數來將輸入層傳輸進來的非線性多維資料映射至輸出層，隱藏層的運作方式根據不同資料型態、預測目標、演算思維會有不同的設置。神經網路內部的神經元數量及層數深度通常由實驗測試決定。

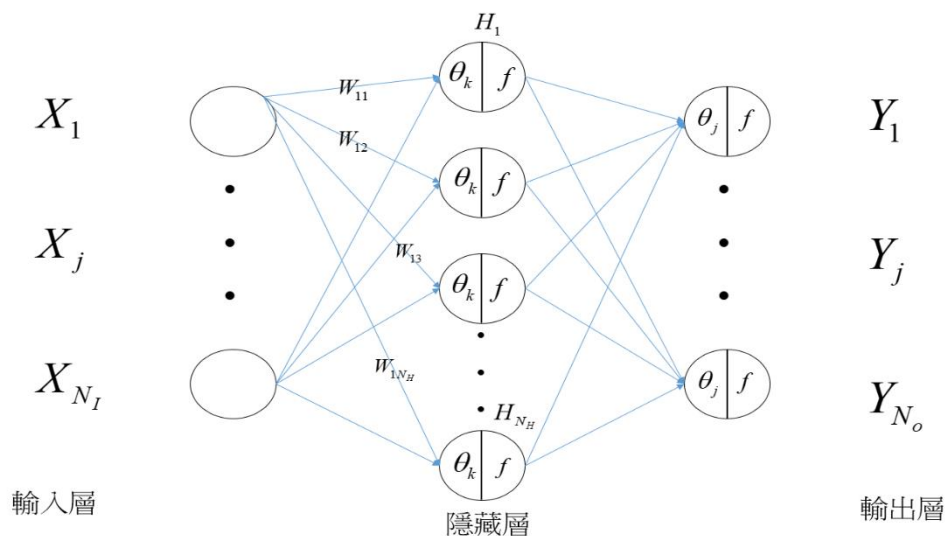


圖 2-1 類神經網路架構示意圖

資料來源：神經網路架構圖-

<https://nccur.lib.nccu.edu.tw/bitstream/140.119/35873/6/25700606.pdf>

根據隱藏層的設計，類神經網路又可分為前饋式類神經網路 (Feedforward Neural Network, FNN) 和遞迴式類神經網路 (Recurrent Neural Network, RNN)。FNN 的資料傳遞方向只有單一方向，不會構

成迴圈，又能依據其複雜度分為單層式 FNN 和多層式 FNN，早期所使用的感知器皆屬於單層式，僅將層次分為輸入層和輸出層，只能處理較不複雜的非線性問題；而多層式意即不同層的神經元互有關連性，能夠更具關連程度給予下一層更具備資料特性之輸出。

多層前饋式類神經網路能夠經由不同層神經元關聯性，將多維的輸入資料映射至輸出值上，因此本計畫採用 FNN 做為崩塌機率預測模型的訓練架構(如圖 2-2)。

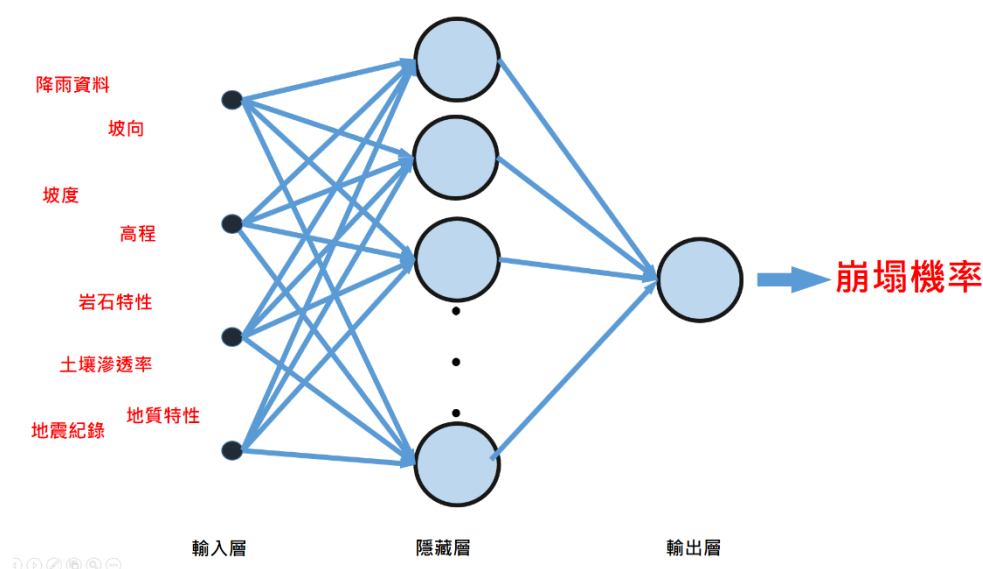


圖 2-2 FNN 模型架構圖

資料來源：本計畫繪製

為了考量多維資料之前後關聯，RNN 中的神經元輸出之資料會回傳至同一層的神經元(圖 2-3)，以此方式上一時刻的資訊回傳至網路中，使網路能考慮前一時刻網路產生之輸出值，該流程使得類神經網路具備記憶性，能儲存並回饋至下一筆的資料處理中，其應用也可以擴展到與時間序列或有前後關聯的預測上。遞迴式類神經網路的應用也不限於任何領域及型態，如於 2001 年，張等人將其用於推估精確度較不易掌握的河川流量推估。於 2017 年，鄭將遞迴類神經網路

應用在複雜、非線性且非穩態的匯率預測、苟等人將其運用在社交網路中消息爆發的傳播過程預測。本計畫中應用之 LSTM 即屬遞迴式類神經網路中的一個分支，其相較於遞迴式類神經網路，LSTM 透過四個閥門控制記憶單元的記憶功能，以解決梯度消失(vanishing gradient problem)以及梯度爆炸(exploding gradient problem)的問題。

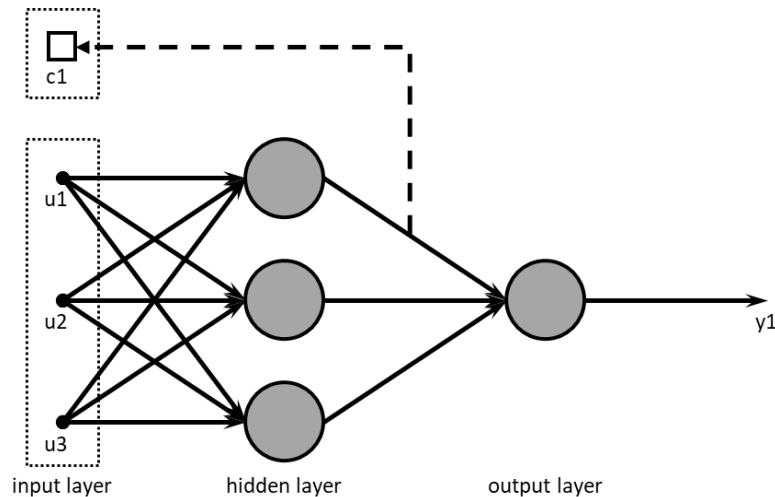


圖 2-3 遞迴式神經網路架構示意圖

資料來源：心理學維基百科

第二節 長短期記憶模型 (Long Short-Term Memory, LSTM)

LSTM 為 RNN 的一種，具備儲存前一次隱藏層輸出值的功能，使得下次輸出值不僅受到當前輸入值影響，更會根據前一次的儲存參數而有所調整，帶著舊有記憶之訊息進入下一層神經元，以此方式不斷遞迴下去(如圖 2-4)。RNN 架構的神經網路由於網路權重的不斷傳遞與相乘，有時容易出現梯度爆炸的問題，使得神經網路運作成效不佳，因此 LSTM 在 RNN 的基礎上增加了三個閥門控制記憶單元，分別為輸入閥(input gate)、輸出閥(output gate)、遺忘閥(forget gate)，輸

入閥決定是否將資料寫入記憶單元，輸出閥控制其他神經元是否從記憶單元中將值讀出，遺忘閥控制過去記憶資料並決定是否保存或刪除。各閥門的激活權重由類神經網路在訓練過程中學習。

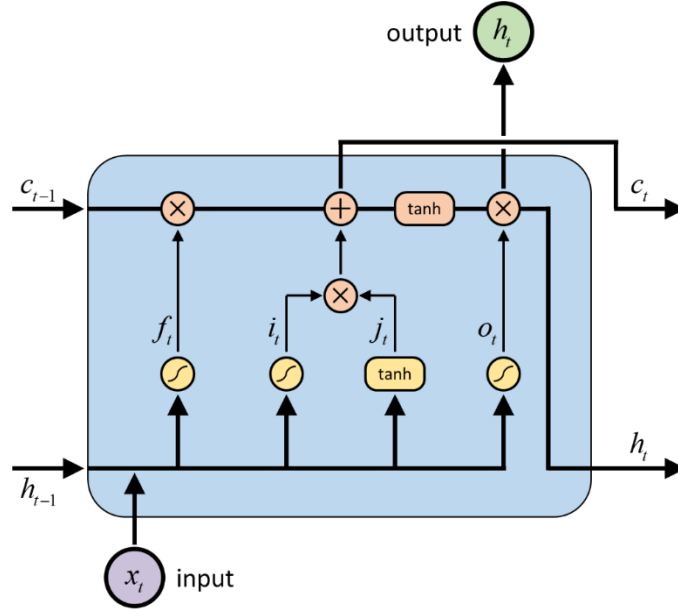


圖 2-4 LSTM 網路架構示意圖

資料來源：Water Quality Prediction Based on LSTM and Attention Mechanism: A Case Study of the Burnett River, Australia

首先，在某個時間步幅 t 下之輸入值會如式 1 中的 $\tanh$ 函數來控制。

$$j_t = \tanh(W_x X_t + W_{hj} h_{t-1} + b_j) \quad (1)$$

式 1 中 j_t 為輸入值通過 $\tanh$ 之結果， X_t 為 t 時刻之輸入值， h_{t-1} 則為 $t-1$ 時刻之短時記憶，也就是 $t-1$ 時刻 LSTM 之輸出值， b_j 為偏差值。而另外三個閥門則是由 sigmoid 函數控制，由 0 或 1 來界定閥門的開啟或是關閉，輸入閥、遺忘閥、輸出閥分別以式 2、3、4 表示。

$$i_t = \text{sigmoid}(W_{xi} X_t + W_{hi} h_{t-1} + b_i) \quad (2)$$

$$f_t = \text{sigmoid}(W_{xf} X_t + W_{hf} h_{t-1} + b_f) \quad (3)$$

$$o_t = \text{sigmoid}(W_{xo}X_t + W_{ho}h_{t-1} + b_o) \quad (4)$$

其中 i_t 、 f_t 、 o_t 分別為輸入值通過輸入閥、遺忘閥、輸出閥之結果。

此外，長時記憶 c 以式 5、6 來表示。

$$c_t = f_t \circ c_{t-1} + i_t \circ j_t \quad (5)$$

$$h_t = \tanh(c_t) \circ o_t \quad (6)$$

其中 c_t 和 c_{t-1} 為 t 時刻和 $t-1$ 時刻之長期記憶，而 $\circ$ 符號代表阿達瑪積，為矩陣運算中的逐項乘積， h_t 則為 LSTM 在 t 時刻之最後輸出值。計算出 c_t 值後，即可當作下一時刻的長時記憶輸入值，且 c_t 值再由 $\tanh$ 函數和輸出閥的控制值 o_t 相乘，即可得到則為 t 時刻輸出值 h_t 。且新的輸出值 h_t 不僅會成為該神經元下一個時刻的輸入值，更會遞迴至同一隱藏層的其他神經元以及下一個隱藏層之神經元，以此方式不斷遞迴下去。

第三章 模型訓練資料庫建立

第一節 選定研究範圍

本計畫在研究範圍上，將針對近 5 年崩塌事件進行選取。位於各個地區的崩塌事件，將提高蒐集資料的多樣性，減少使用上的地區性。若能找出時間序列資料與該地區環境因子的崩塌關係，將能提前預防並提供更快速的救災。基於上述原由，本研究團隊選取臺灣北、中、南、東各地區崩塌相對密集的區域作為本計畫之研究範圍，透過林務局農林航測所、水土保持局資料開放平台、國土測繪中心，獲取選定區域圖層以及周邊相關地質、水文等地理資訊圖層，以獲取訓練模型所需要之地文資料。透過 CODiS 來獲取得各個測站的降雨資料，以一小时之時雨量資料為單位整理成降雨時序資料，並結合由中央氣象局地震測報中心之測站所得地震資料組成之地震時序資料。

第二節 基本資料庫說明

為了進行人工智慧分析，需先建立基本資料庫，本計畫將結合環境因子和時間序列因子進行坡地崩塌機率評估，分別進行兩者的基本資料收集，以作為後續模型預測崩塌潛在機率所使用的輸入資料。

環境因子的部分，蒐集相關地質、水文等地理資訊、集水區基本圖資等資料，並使用內政部之數值高程模型，結合 QGIS 等地理資訊軟體(圖 3-1)，獲取高程、計算坡度等各項地形參數，將其整理成 20mX20m 網格資料庫，連結研究範圍內之各項屬性因子，建置環境因子資料，其中包含坡度、高程、坡向、植被覆蓋狀況等地理特質等。崩塌歷史事件資料採用事件型崩塌目錄配合全球災害事件簿之相關

調查報告做為參考資料。

時間序列因子採用長短期記憶模型來進行估算，採用該區域五年來之地震歷史資料，將地震歷時經過資料處理後已周為單位組成 72 周的時間序列，並以此當 LSTM 的訓練及驗證資料，上述之資料即是地震時間序列資料；降雨資料使用來自 CODiS 之月資料報表，根據每一筆崩塌事件發生時間向前擷取 216 小時(9 天)之雨量資料做為該崩塌事件之相應總降雨時間序列資料，並以 72 小時的降雨時間做為 LSTM 的訓練及驗證資料。在崩塌機率預測時，亦是使用當下之前 72 小時雨量資料做為時間因子判定的時序資料。為了將時間序列資料之特徵轉換成利於模型預測崩塌機率之數值特性，需將時間序列資料解算成時間序列因子，本計畫使用長短期記憶模型 LSTM 來進行因子解算。

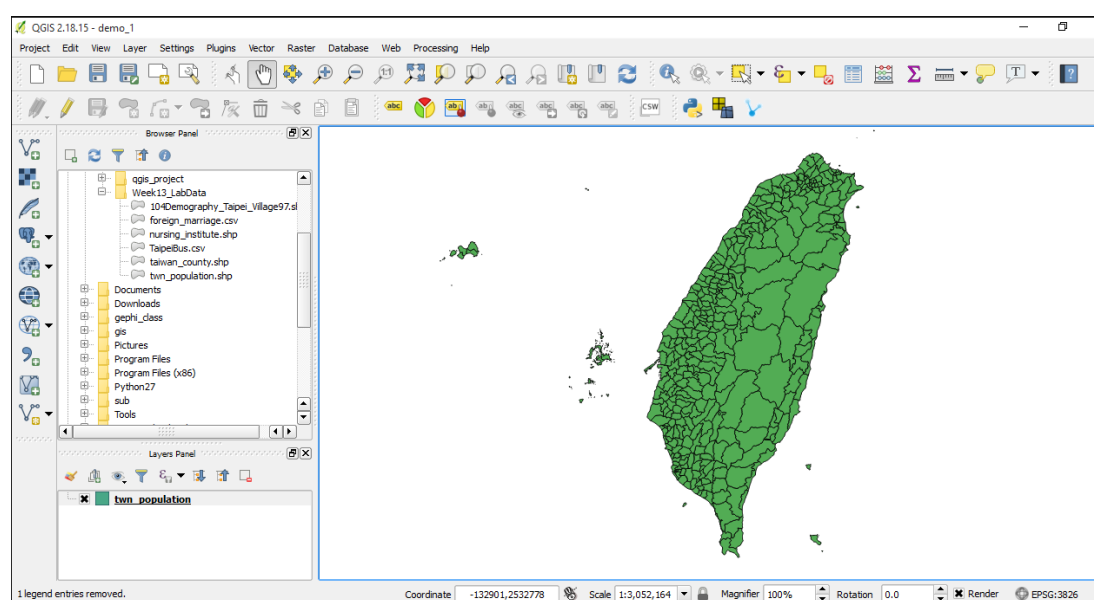


圖 3-1 QGIS 操作介面示意圖

資料來源：本計畫繪製

第三節 資料量化與因子解算

崩塌潛在機率預測模型需接收時間序列因子和環境因子這兩類數值化數據，才能根據模型權重進行機率預測。時間序列因子之解算需使用降雨資料和地震紀錄資料兩項資料，並依照如下步驟進行整理以及解算。

一、降雨資料

- (一) 將待測時間段前72小時之時雨量資料(圖3-2)整理為時間序列資料。
- (二) 將降雨時間序列做為LSTM之輸入資料，並以崩塌事件紀錄之崩塌地區當作訓練標籤進行模型訓練。
- (三) 利用降雨因子值設定方法(如圖3-3)給定LSTM訓練資料之因子值。
- (四) 對模型測試結果以及模型訓練收斂性進行評估，確認模型解算出之時間序列因子為有效後，完成時間序列因子解算模型的建置。
- (五) 模型訓練完成後，將選定地區之崩塌前72小時降雨序列資放入模型，所得出之結果作為後續崩塌機率預測模型之輸入資料。

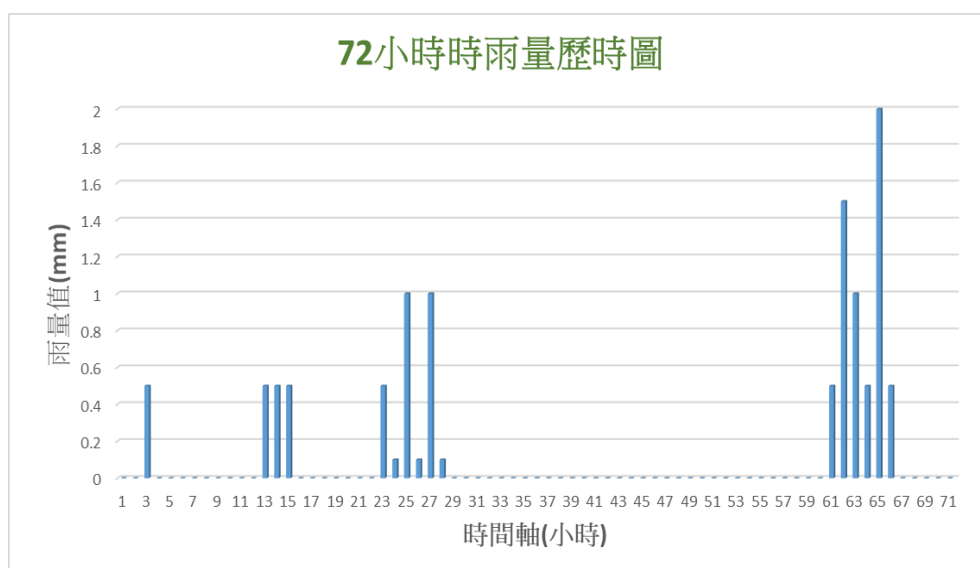


圖 3-2 72 小時之時雨量資料示意圖

資料來源：本計畫繪製

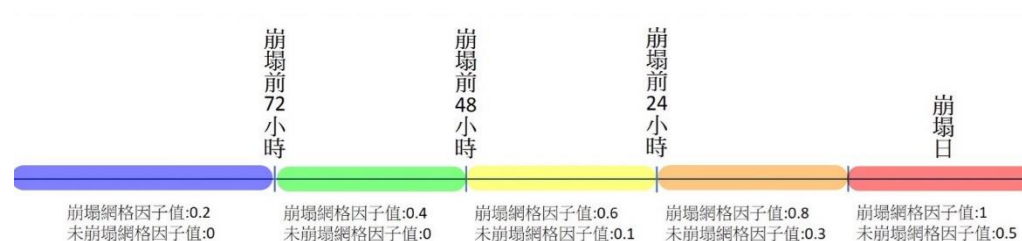


圖 3-3 降雨因子值設定方法

資料來源：本計畫繪製

二、地震資料

- (一) 將待測時間段前72周之時地震強度資料(圖3-4)整理為時間序列資料。
- (二) 將地震強度時間序列做為LSTM之輸入資料，並以崩塌事件紀錄之崩塌地區當作訓練標籤進行模型訓練。
- (三) 利用地震強度因子值設定方法(如圖3-5)給定LSTM訓練資料之因子值。

- (四) 對模型測試結果以及模型訓練收斂性進行評估，確認模型解算出之時間序列因子為有效後，完成時間序列因子解算模型的建置。
- (五) 模型訓練完成後，將選定地區之崩塌前72周地震強度序列資料放入模型，所得出之結果作為後續崩塌機率預測模型之輸入資料。

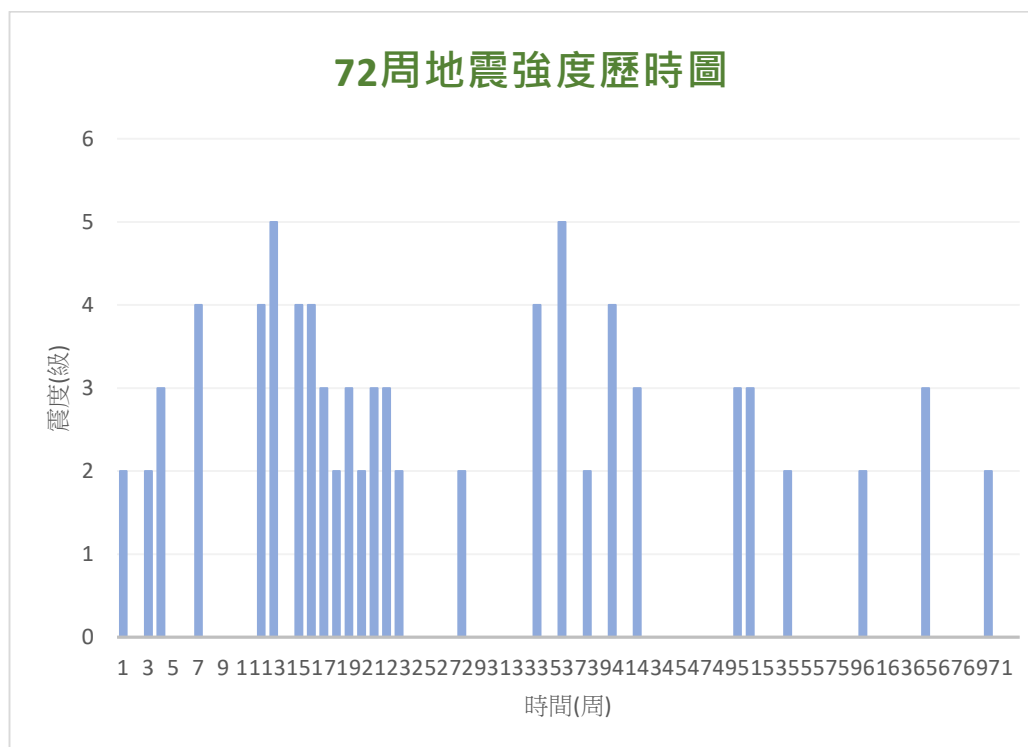


圖 3-4 地震歷史紀錄資料示意圖

資料來源：本計畫繪製

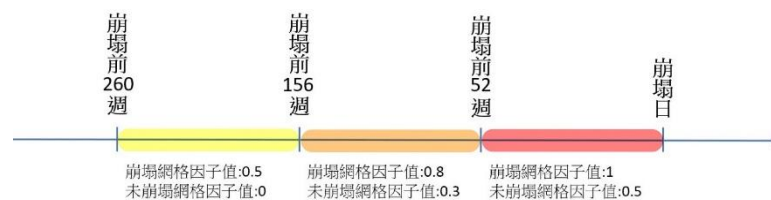


圖 3-5 震強度因子值設定方法

資料來源：本計畫繪製

環境因子之組成包含坡向、坡度、高程、土壤滲透率及植被狀況等地文資料。由於各項資料之量級大小具有差異性，因此需先對資料數據進行標準歸一化，使得各屬性間不會有過大的數值量級差距。植被狀況之數值量化常見方式有兩大種類，一為計算區域內的植被覆蓋率，計算方式該區域之區域植物和該區域面積之比值；另一種類為植生指標，植生指標是兩個或多個可見光與近紅外波長範圍內的目標物表面反射率的組合計算，目的為增量植被的某一特徵或細節。常見的指標有 NDVI(常態化差異植生指標)、EVI(增強植被指標)、RVI(紅光比植生指標)等等，而這次計畫選擇使用 NDVI 做為環境因子。

第四節 DEM 模型 20m x 20m 網格

數字高程模型(Digital Elevation Model, DEM)是通過有限的地形高程數據實現對地面地形的數字化模擬，使地形表面形態能夠以數字化表達，為一種以一組有序數值形式表示地面高程的一種實體地面模型。為了將資料有效量化以利於人工智慧模型進行機率預測，本計畫將前面小節提及之因子與 DEM 做結合，形成以網格為單位的資料集合。網格建置範圍是以事件型崩塌目錄類所標柱的崩塌範圍為基準，向外擴展一格做為網格範圍(如圖 3-6)。本計畫預計先以 20m x 20m (如圖 3-7)的網格大小做為分析最小單元，並在測試模型的過程中根據預測結果調整網格大小。透過預測結果的誤差分析，找到一個最適當的網格維度做為資料集合的大小。

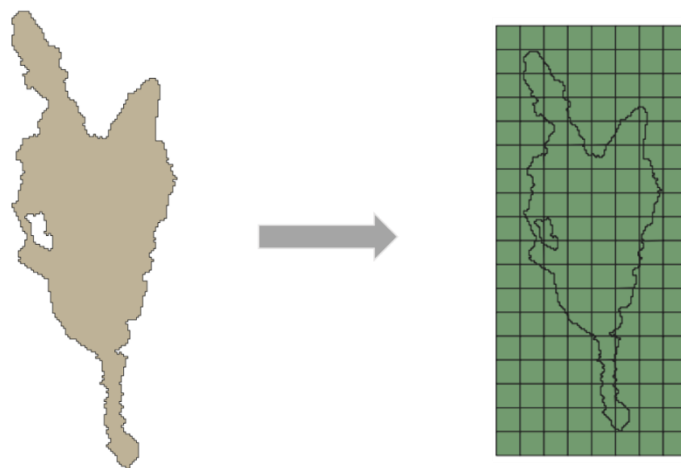


圖 3-6 網格建置示意圖

資料來源：本計畫繪製

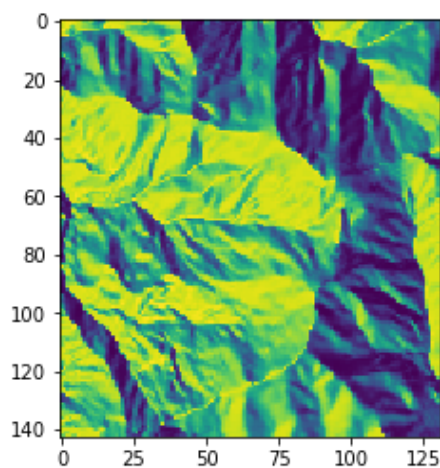


圖 3-7 20m x 20m 網格示意圖

資料來源：本計畫繪製

第五節 網格資料庫建置

崩塌機率預測模型和時間序列因子解算模型都需要訓練樣本資料做為輸入資料，因此我們需預先建立一以網格維度大小為單位之資料庫，將收集而來之資料和網格單位做有效的結合。各資料初步預想之結合方式如下所列，其步驟以及方法將會在之後的結果評估中做適

當調整以利於模型訓練之收斂性與成效。

一、降雨資料：收集距網格中心最近之三個雨量降雨測站之資料。

二、地震歷史紀錄：使用距網格中心最近且資料齊全之地震測站資料，並假設網格所接收到的地震震度與測站相同。

三、坡向：選取網格範圍內之九個可能坡向之比例中最大者做為該往個的代表坡向(如圖3-8)，九個可能坡向包含平地、東、西、南、北、東北、東南、西北、西南。根據內政部所提供的網格數值模型地形資料，內容包括透過顏色表達網格地形分布的GRD檔案及單純儲存崩塌文字資料像是平面座標系、圖幅西南偶E座標、圖幅西南偶N座標等等的HDR檔案，將GRD檔匯入QGIS後儲存成tiff檔，並使用內建的程式將tiff檔轉成shp檔後即可利用QGIS內建的坡向分析程式可以將用顏色表達的地形分布圖以數值化的方式呈現。

坡向類別	格點百分比	格點面積 (平方公尺)
平地	0.00000000	0.0
北	0.00000000	0.0
東北	0.00000000	0.0
東	0.00000000	0.0
東南	0.00000000	0.0
南	0.00000000	0.0
西南	0.00000000	0.0
西	0.00000000	0.0
西北	1.00000000	3600.0

圖 3-8 坡向資料

資料來源：本計畫繪製

四、坡度：根據範圍內的坡度分級(圖3-9)和面積占比之乘積決定該網格範圍之平均坡度。根據內政部所提供的網格數值模型地形資料，內容包括透過顏色表達網格地形分布的GRD檔案及單純儲存崩塌文字資料像是平面座標系、圖幅西南偶E座標、圖幅西南偶N座標等等的HDR檔案，將GRD檔匯入QGIS後儲存成tiff檔，並使用內建的程式將tiff檔轉成shp檔後即可利用QGIS內建的坡度分析程式可以將用顏色表達的地形分布圖以數值化的方始呈現。

坡度	格點百分比	格點面積 (平方公尺)
一級坡	0.0	0.0
一級坡	0.0	0.0
三級坡	0.1875	1200.0
四級坡	0.1875	1200.0
五級坡	0.25	1600.0
六級坡	0.375	2400.0
七級坡	0.0	0.0

圖 3-9 坡度資料

資料來源：本計畫繪製

五、高程：網格內之所有高程點之高程值平均做為網格高程(圖3-10)。根據內政部所提供的網格數值模型地形資料，內容包括透過顏色表達網格地形分布的GRD檔案及單純儲存崩塌文字資料像是平面座標系、圖幅西南偶E座標、圖幅西南偶N座標等等的HDR檔案，將GRD檔匯入QGIS後儲存成tiff檔，並使用內建的程式將tiff檔轉成shp檔後即可直接查看網格高程。

#	X	Y	Z
1	240918.794782	2679330.200675	994.423328
2	240881.738297	2679312.886241	988.456267
3	240883.892369	2679273.873819	1001.108777
4	240947.097293	2679289.004315	1027.32126
5	240927.461434	2679252.173044	1007.581902

圖 3-10 高程資料

資料來源：本計畫繪製

六、植被狀況：以網格範圍為單位進行植被狀況的量化結果，得出之結果代表該網格之植被狀況屬性，使用EO Browser中的Sentinel-2衛星資料(如圖3-11)，將從網站下載的NDVI網格資料匯入QGIS。

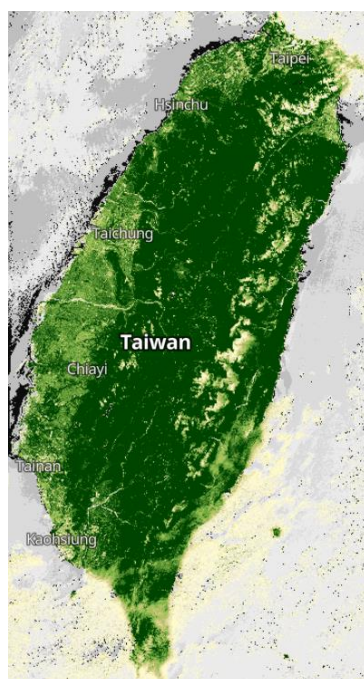


圖 3-11 Sentinel-2 衛星之 NDVI 圖

資料來源：EO Browser

第六節 崩塌坡地資料庫建立成果

本計畫目標為建立一套應用於臺灣地區的崩塌預警系統，因此在資料庫建立上，需考量臺灣地區各地特性，在 QGIS 中觀察這 10 年間的崩塌資料後，決定選花蓮、新北、南投、台中、嘉義為建置崩塌網格資料庫地區，目前總共收集的網格資料數目為 1273，崩塌網格及未崩塌網格數量分別 606 及 667 為。以下以花蓮資料建置資料庫過程為例，總共有 144 格網格資料，分別為 66 格崩塌網格及 78 格未崩塌網格，選定區域為帕托魯山附近的崩塌地區(如圖 3-12)，崩塌時間約為 2017 年 6 月 2 號至 4 號之間。

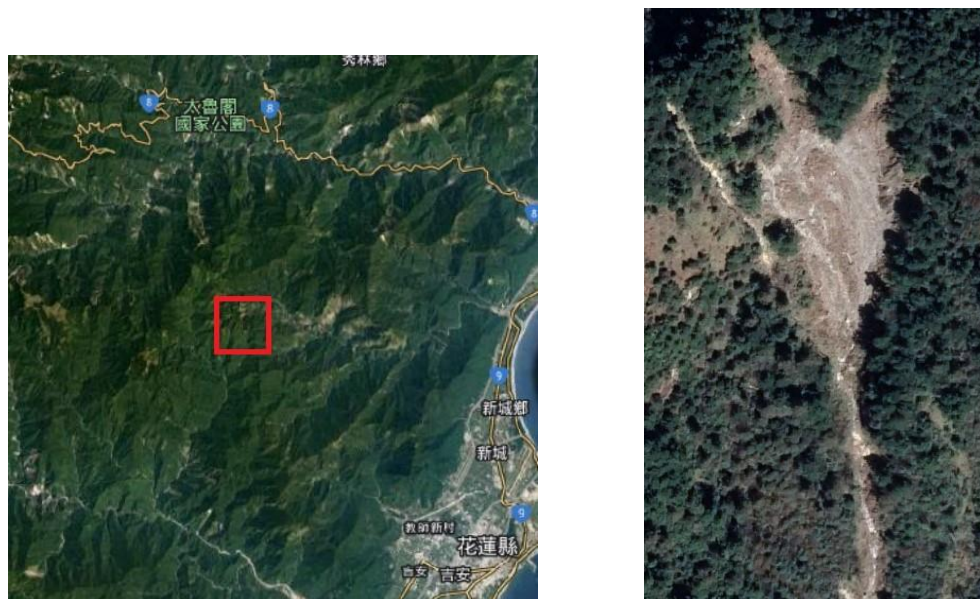


圖 3-12 花蓮縣帕托魯山附近的崩塌地區示意圖

資料來源：本計畫繪製

從數位發展部的政府資料開放平台下載事件型崩塌目錄，內容包含 20 公尺網格數值模型資料總共 18,733 格網格，利用 QGIS 內建的網格繪製程式設定邊界及網格長度和寬度分別為 20 公尺繪製網格，

找到崩塌區域及網格重疊部分後將其他網格刪除，只留下崩塌附近網格(如圖 3-13)，作為環境因子的蒐集及分析對象，各項環境因子蒐集後之資料分述如下。

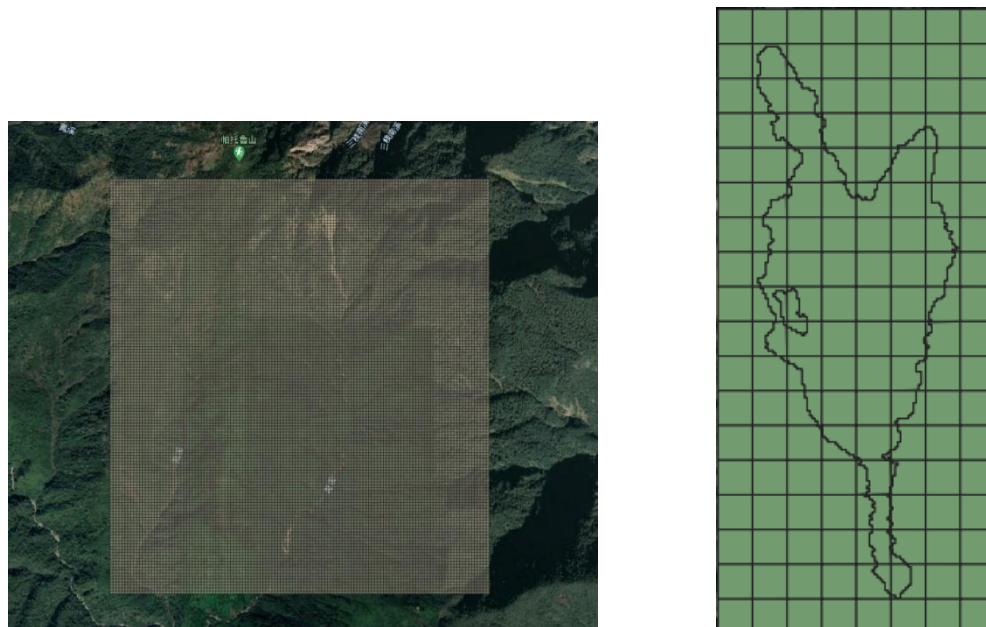


圖 3-13 花蓮縣帕托魯山附近的崩塌範圍設定示意圖

資料來源：本計畫繪製

一、降雨資料

作為時間序列因子的一部分，透過觀測資料查訊系統來獲取資料並將其整理成崩塌前 72 小時之降雨時間序列資料，透過測站資料得出座標位置，配合網格座標即可找出距離單一網格最近的三個測站，並利用距離加權的方法將降雨量轉換成更接近實際網格位置之加權降雨量，以供後續長短期記憶模型模型 (Long Short-Term Memory, LSTM) 解算時間序列因子，這筆花蓮崩塌所發生的時間約為 2017 年 6 月 2 號至 4 號，所以蒐集的地震資料時間是從 2017 年 5 月 27 號至 2017 年 6 月 4 號，以 1 小時為一個區間記錄資料，總共 169 筆資料，最後做出網格降雨時間序列資料(如表 3-1)，第 1 格網格降

雨時間序列圖(如圖 3-14)及第 20 格網格降雨時間序列圖(如圖 3-15)如下。

表 3-1 網格降雨時間序列資料

小時 網格	1	2	3	4	5
1	0.343523	0	0.843523	0.343523	0
20	0.343642	0	0.843642	0.343642	0
小時 網格	6	7	8	9	10
1	0.199310	0.199310	0	0	0
20	0.199235	0.199235	0	0	0

資料來源：本計畫整理

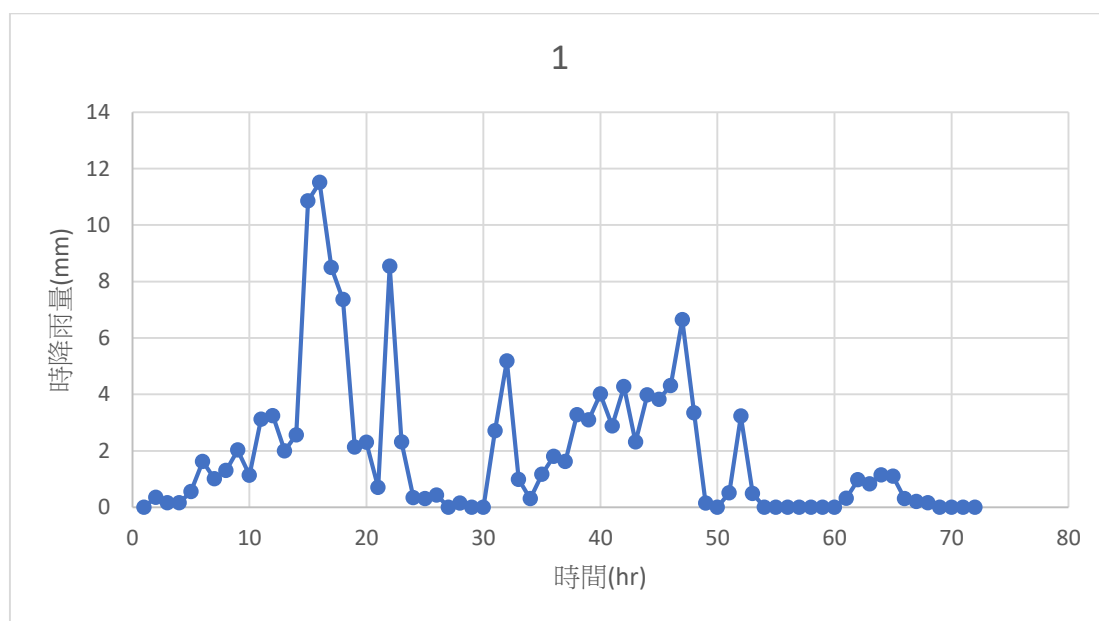


圖 3-14 花蓮 6/2~6/4 的第 1 格網格(未崩塌)降雨時間序列

資料來源：本計畫繪製

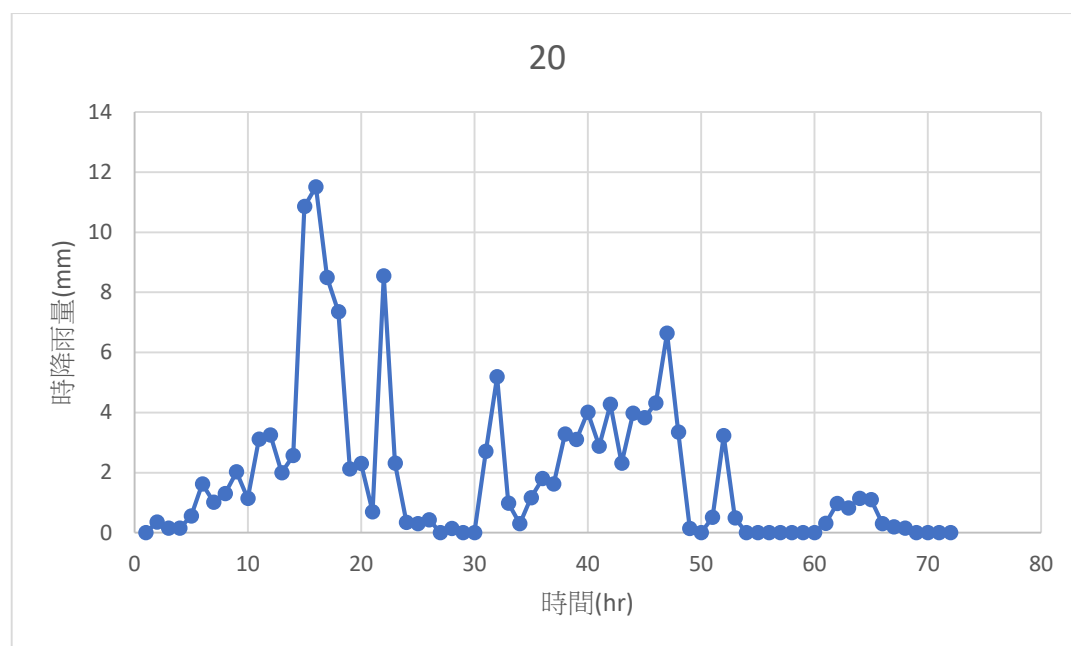


圖 3-15 花蓮 6/2~6/4 的第 20 格網格(崩塌)降雨時間序列

資料來源：本計畫繪製

二、地震紀錄

作為時間序列因子的一部分，由中央氣象局地震測報中心(如圖 3-16)查詢選定地區距離崩塌中心最近且資料最齊全之地震測站，並取得崩塌發生前 5 年之地震震度資料，最後將其量化成時間長度相同的 72 週地震歷史歷時資料，以供後續 LSTM 解算時間序列因子，這筆花蓮崩塌所發生的時間約為 2017 年 6 月 2 號至 4 號，所以蒐集的地震資料時間是從 2012 年 6 月 2 號至 2017 年 6 月 4 號，以 4 週為一個區間記錄資料，總共 445 筆資料，因時間是以周為單位，所以只會記錄一週內最高震度，最後做出網格地震強度時間序列資料(如表 3-2)，第 1 格網格地震震度時間序列圖(如圖 3-16)及第 20 格網格地震震度時間序列圖(如圖 3-17)如下。

銅門國小

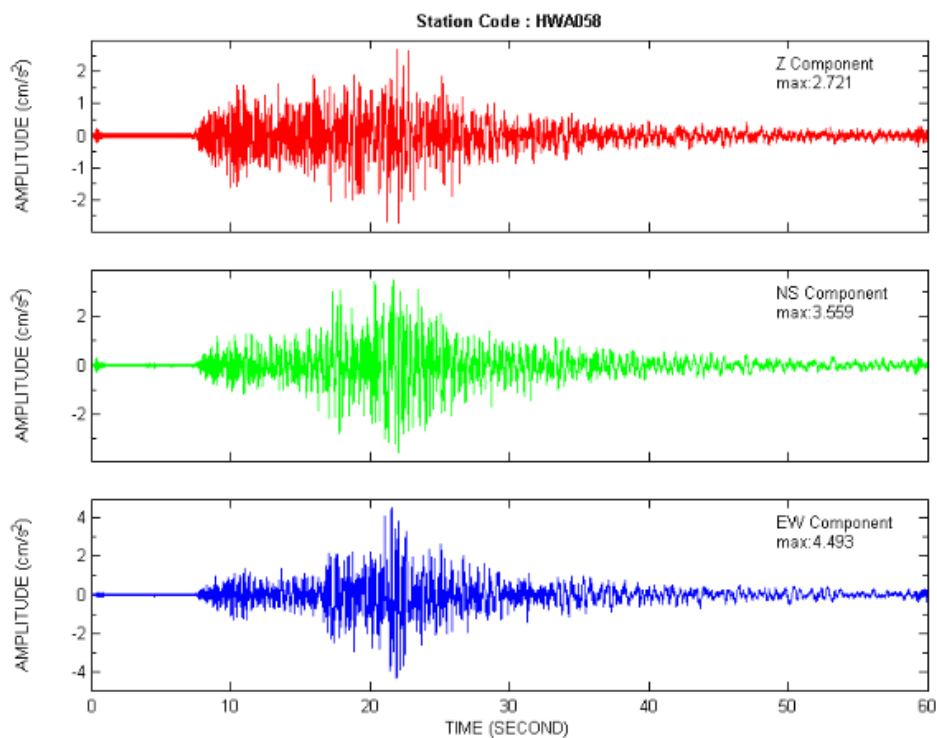


圖 3-16 地震測報中心的震度資料

資料來源：地震測報中心

表 3-2 網格地震震度時間序列資料

震度 表格	1	2	3	4	5	6	7	8	9	10
1	2	0	2	3	0	0	4	0	0	0
20	2	0	2	3	0	0	4	0	0	0

資料來源：本計畫整理

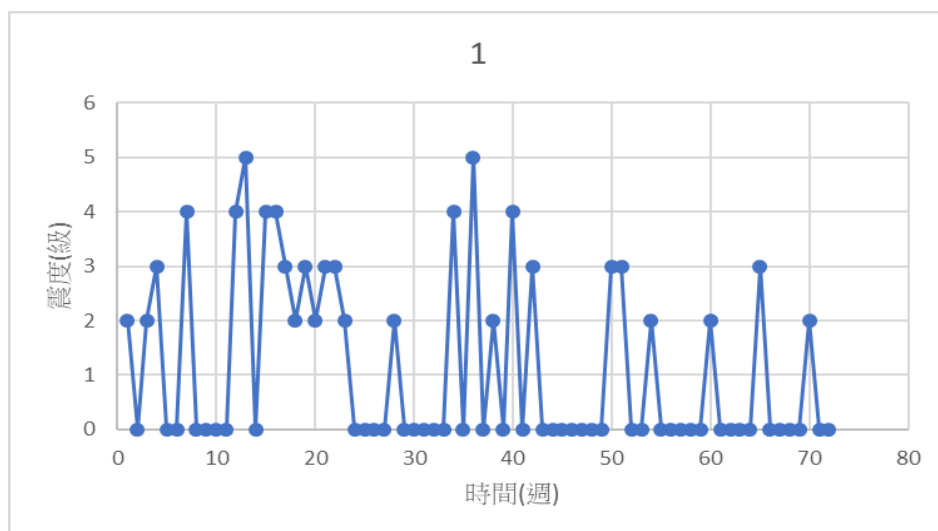


圖 3-17 花蓮 6/2~6/4 的第 1 格網格(未崩塌)地震震度時間序列

資料來源：本計畫繪製

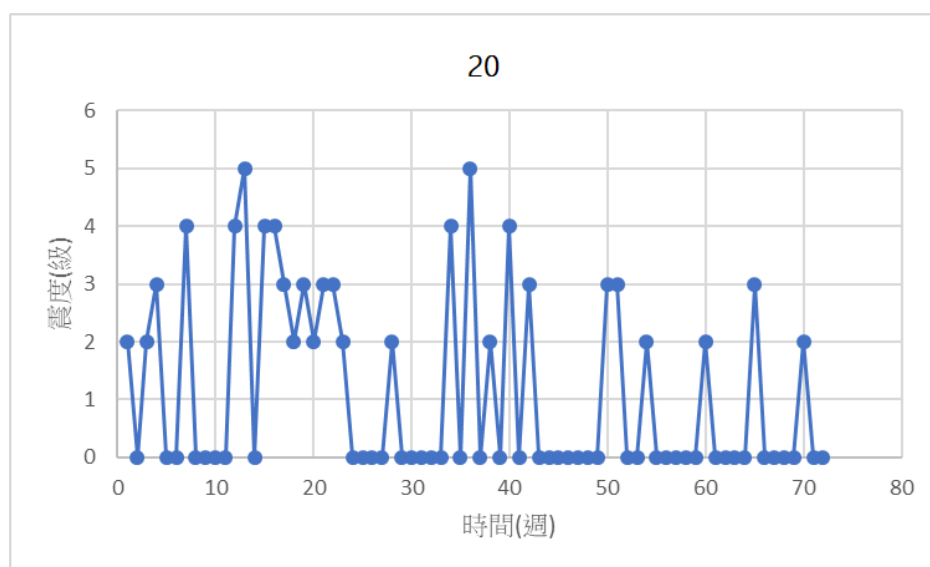


圖 3-18 花蓮 6/2~6/4 的第 20 格網格(崩塌)地震震度時間序列

資料來源：本計畫繪製

三、坡向

為考量的環境因子其中之一，根據上節所述步驟再透過地圖顏色調整做出臺灣坡向分布(如圖 3-19)，以取得目標區域的坡向資料，並彙整入資料庫和其他地文資料一起量化成環境因子，透過觀察發現經

過分析出來的網格數目減少為 18,189 格，因為 QGIS 內建的坡向分析方法為利用周邊 8 格網格推斷出中間網格的坡向數據，所以邊界網格會消失，花蓮地區網格坡向圖(如圖 3-20)及數值化呈現的坡向表格(如表 3-3)如下。

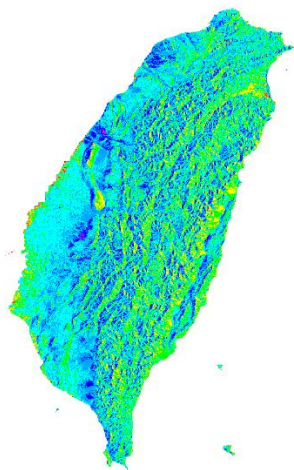


圖 3-19 臺灣坡向分布圖

資料來源：本計畫繪製

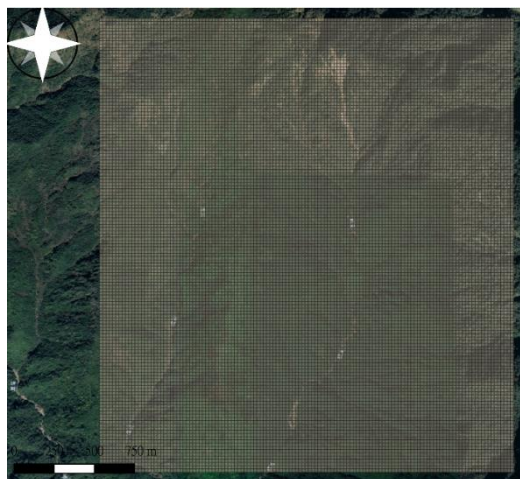


圖 3-20 花蓮地區網格坡向圖

資料來源：本計畫繪製

表 3-3 數值化呈現的網格坡向

網格	坡向	網格	坡向
1	338.9377	7	126.3251
2	355.1035	8	124.5929
3	143.5009	9	130.9311
4	143.1634	10	135.1532
5	133.4126	11	135.6270
6	130.0665		

資料來源：本計畫整理

四、坡度

為考量的環境因子其中之一，根據上節所述步驟即可將資料以網格(如圖 3-21)及數值化的方始呈現(如表 3-4)，以取得目標區域的坡度資料，並彙整入資料庫和其他地文資料一起量化成環境因子，透過觀

察發現經過分析出來的網格數目減少為 18,189 格，因為 QGIS 內建的坡度分析方法為利用周邊 8 格網格推斷出中間網格的坡度數據，所以邊界網格會消失。



圖 3-21 花蓮地區網格坡度圖

資料來源：本計畫繪製

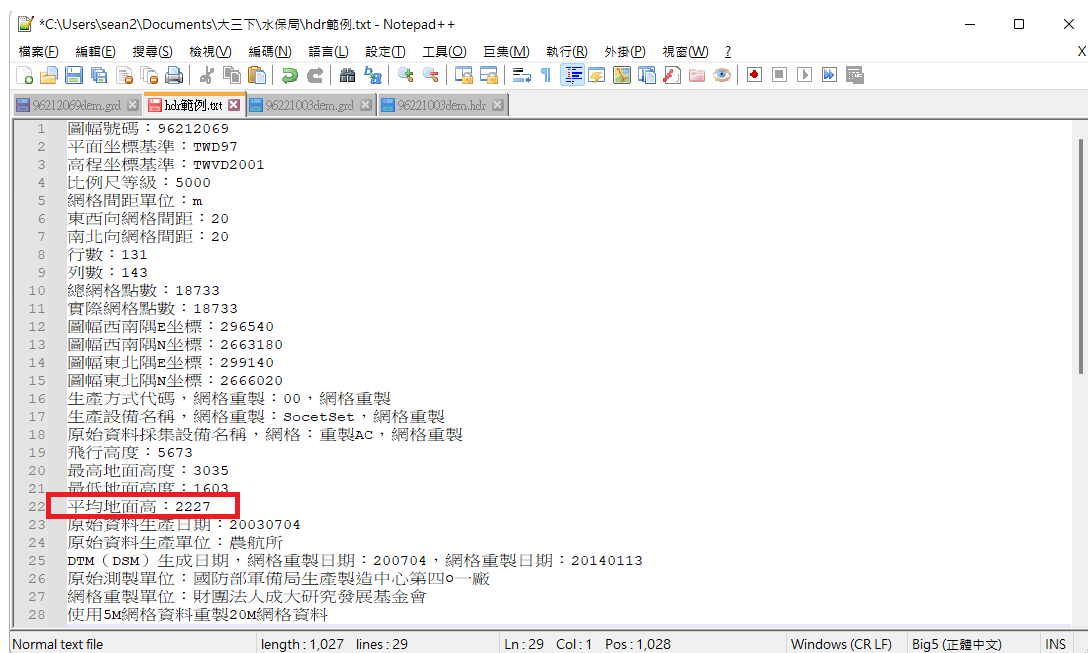
表 3-4 數值化呈現的網格坡度

網格	坡度	網格	坡度
1	40.8831	7	48.4582
2	19.8152	8	49.1925
3	34.7663	9	53.0935
4	49.5896	10	56.6320
5	48.2407	11	56.6349
6	48.5838		

資料來源：本計畫整理

五、高程

為考量的環境因子其中之一，網格中心高程數值包含在 HDR 檔案中(如圖 3-22)，也可以根據上節步驟將 GRD 檔中用顏色表達的地形分布圖以數值化的方始呈現(如表 3-5)，並製作出台灣高程陰影圖(如圖 3-23)，以取得目標區域的高程資料，並彙整入資料庫和其他地文資料一起量化成環境因子，本計劃高程數據是利用 QGIS 轉換所得出之數值。



資料來源：本計畫繪製

表 3-5 數值化呈現的網格高程

網格	高程	網格	高程
1	2072.03	7	2078.19
2	2056.92	8	2099.91
3	2054.91	9	2116.83
4	2057.12	10	2137.57
5	2054.53	11	2165.19
6	2065.42		

資料來源：本計畫整理



圖 3-23 臺灣高程陰影圖

資料來源：本計畫繪製

六、植被狀況

為考量的環境因子其中之一，根據上訴步驟下載檔案(如圖 3-24)，此衛星為 10m x 10m 的網格，所以需轉換為本計畫所設計的 20m x 20m 網格，最後即可得出 20m x 20m 之網格 NDVI 值(如表 3-6)。

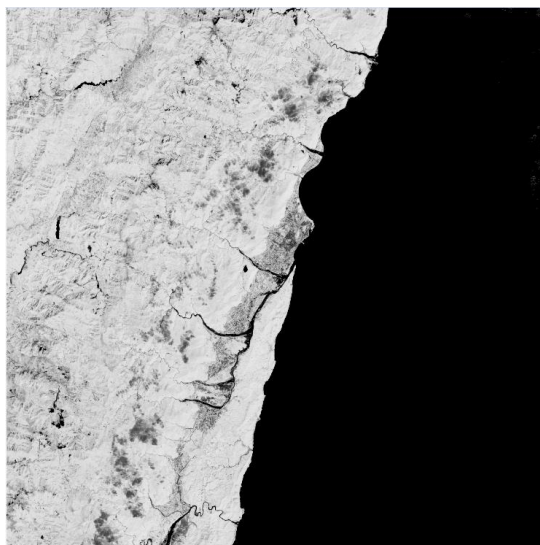


圖 3-24 花蓮 NDVI 圖

資料來源：本計畫繪製

表 3-6 網格 NDVI 值

網格	高程	網格	高程
1	0.531618	7	0.695733
2	0.669257	8	0.687873
3	0.684444	9	0.627608
4	0.692372	10	0.685308
5	0.625902	11	0.597213
6	0.636377		

資料來源：本計畫整理

第四章 模型測試與訓練

第一節 模型建置流程

使用 LSTM 模型對時間序資料進行解算並得出時間序列因子，配合地文資料整理而成之環境因子，最後利用前饋式類神經網路來計算每格大小為 20m*20m 之網格崩塌機率。

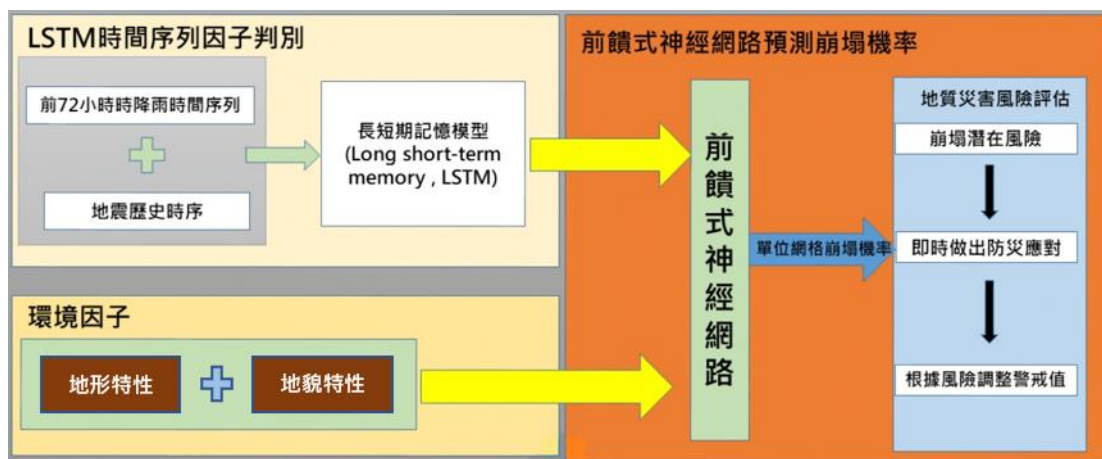


圖 4-1 模型流程圖

資料來源：本計畫繪製

第二節 LSTM 網路建置

本計畫預計使用 Tensorflow 套件來進行類神經網路架構的相關建置，Tensorflow 為 Google 所開發並開源以及維護之深度學習框架，該框架由 Python 語言進行編寫，使用上不僅介面簡單、擴充性高，能夠降低使用者建置深度學習網路的門檻，並且能以較為簡潔高效的程式碼進行機器學習模型的建立、測試、評估與預測。

時間序列資料對於模型而言是維度非常大的資料，因為一筆時間序列中根據頻率或是區間劃分可能會有數十甚至數百個區間輸入至模型中，對於模型訓練而言若尺度太大則訓練效果和效率會大幅降低，

造成訓練時間過長或是預測結果不佳的情況。因此在資料輸入前須先進行自然語言處理中較常使用的嵌入層(Embedding Layer)，其概念為把維度為所有詞語數量的高維空間嵌入到維度低的連續向量空間，利用低維度的連續向量，值不一定為 0 或 1，去表示原本較長之 0、1 數列，不僅是詞語，只要能夠透過嵌入層將數字序列轉換為向量皆可使用。

LSTM 網路的架構圖(圖 4-2)需經過實驗以及預測結果的驗證評估決定，嵌入層能使訓練資料的維度降低，但若超參數設定不恰當，在資料壓縮的過程中亦容易出現資料特徵遺失之情形。嵌入層的後方連接 LSTM 層，LSTM 層的神經元數將對應嵌入層之維度長度，並且將層中之記憶單元內將輸入值、過去的輸出值與神經元全連接。隨後再添加兩層隱藏層，神經元數分別為 80 和 20，此二層隱藏層使用之激發函數皆為 ReLU(圖 4-3)。最後，因為本計畫欲將 LSTM 應用在降雨資料與地震紀錄時間序列之因子解算，其標籤值會根據崩塌發生相關紀錄進行 0~1 的設置，因此輸出層尺度為 1，並使用 sigmoid 函數(圖 4-4)將輸出值維持在 0~1 之間。在各層中加入一個 dropout 層，其概念是隨機將一些神經元關閉，避免神經元之間的過度依賴關係，防止過擬合(Over fitting)現象。由於模型的預測目標為數值 0~1 之因子值，因此使用的損失函數為 mean-square error (MSE)，其計算如下式所示：

$$MSE_Loss = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (7)$$

其中， y_i 為目標真正的實際值， $\hat{y}_i$ 為類神經網路的輸出值，N 則為測試樣本筆數。

Model: "sequential_1"

Layer (type)	Output Shape	Param #
embedding_1 (Embedding)	(None, 151, 10)	4020
dropout_1 (Dropout)	(None, 151, 10)	0
lstm_1 (LSTM)	(None, 10)	840
dense_1 (Dense)	(None, 80)	880
dropout_2 (Dropout)	(None, 80)	0
dense_2 (Dense)	(None, 20)	1620
dropout_3 (Dropout)	(None, 20)	0
dense_3 (Dense)	(None, 1)	21
Total params: 7,381		
Trainable params: 7,381		
Non-trainable params: 0		

圖 4-2 LSTM 模型架構概述示意圖

資料來源：本計畫繪製

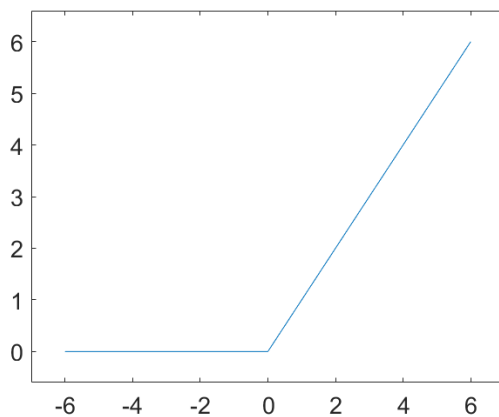


圖 4-3 ReLU 激發函數示意圖

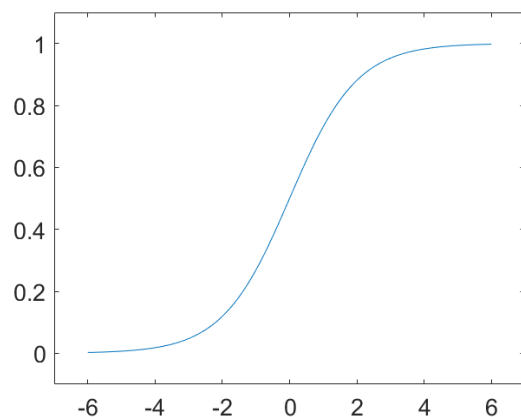


圖 4-4 sigmoid 激發函數示意圖

資料來源：本計畫繪製

第三節 網路訓練和驗證

在 LSTM 的訓練與測試中，將資料庫中不同時段的降雨時間序

列和地震紀錄時序列做為樣本資料，隨機選取樣本資料筆數的 80%做為訓練樣本，其餘約 20%樣本做為測試樣本。每筆資料都包含降雨時間序列和地震紀錄時間序列，以及與之對應的崩塌事件類別標籤。本計畫之標籤設定為當該時段之網格有崩塌事件發生時，將標籤值設為 1，該時段之網格未有崩塌事件發生時將標籤值設為 0。訓練樣本在類神經網路中經由嵌入層轉換與 LSTM 網路計算後，進行多次訓練與測試，並根據訓練結果產生 epoch-loss 曲線圖。產生曲線圖後，需要對曲線趨勢進行評估。若曲線趨勢逐漸隨著 epoch 數收斂，表示模型在訓練過程中，模型的權重分配和轉換方式能夠有效的將輸入資料映射至輸出資料上。曲線圖又分為訓練曲線圖(圖 4-5)和驗證曲線圖(圖 4-6)，後者為模型在訓練過程中使用從訓練資料集獨立分割出來的驗證資料集做為輸入資料所得到的結果，用以判別訓練過程是否出現過擬合情形。過擬合情形表示模型之複雜度超過實際映射該有之複雜度，使得模型的映射行為受到一些資料集中相對極端的資料點干擾，使得模型最終的預測成效不佳。若訓練曲線圖趨勢收斂但驗證曲線圖趨勢沒有隨之收斂，表示可能出現過擬合情形，會產生預測結果不佳的情況。

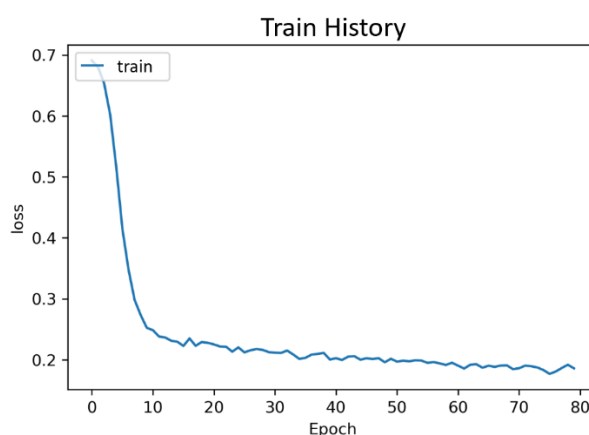


圖 4-5 訓練曲線圖收斂示意圖

資料來源：本計畫繪製

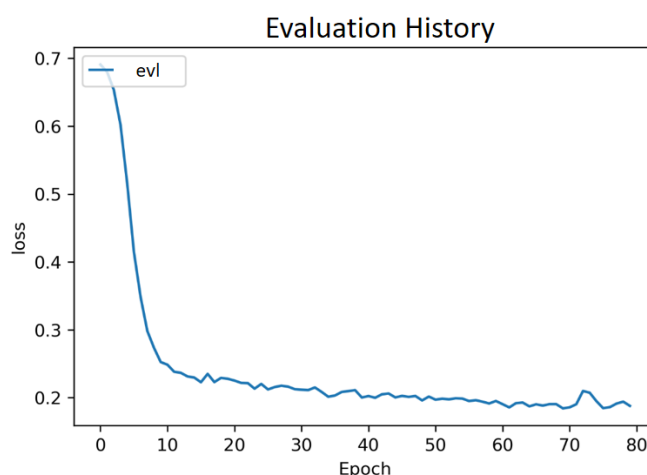


圖 4-6 驗證曲線圖收斂示意圖

資料來源：本計畫繪製

第四節 網路測試以及優化

當網路達到指定的訓練迭代數(epoch)後，訓練便會結束，若訓練曲線圖之趨勢未呈現出收斂性，則需要進行以下幾種調整：

- 一、調整模型訓練超參數，其中包含迭代數(epoch)、學習率(learning rate)、激活函數型別的設置、神經元數量、隱藏層數量等等。
- 二、調整嵌入層之維度轉換之長度，嘗試不同之向量長度檢驗收斂性。
- 三、調整輸入之時間序列資料之取樣頻率，改變時間序列之區間數。
- 四、調整時間序列之擷取以及時間序列之種類，調整資料之量化方式。
- 五、若為過擬合情形，增加訓練資料量來控制方差或是添加一些限制條件式來降低模型複雜度。

當模型訓練完畢，曲線圖表現出收斂性且未出現過擬合情形，表示網路模型在訓練階段表現良好，餘下的 20%測試資料集便能夠直接

做為模型的輸入資料並且透過模型的映射得到預測結果。將預測結果和正確答案之標籤值進行誤差分析，透過誤差結果來評估該模型是否真正的表現良好。若誤差過大，表示可能在此模型有以下幾種情形發生：

- 一、訓練階段雖然曲線趨勢呈現收斂，但仍未達到完全收斂的階段，使得模型表現仍未提升至預期最佳效果，建議提高訓練 epoch 數。
- 二、訓練階段之訓練集和驗證集分割過程產生錯誤，使得訓練階段之過擬合情形並未表現在曲線圖上，造成後續測試階段成效不佳。
- 三、分割訓練集資料和測試資料集時未正確隨機分配，造成分割時將資料屬性高度相似的資料分配在同一集合中。

當模型測試情況良好，我們便可以根據解算出的時間序列因子結合環境因子進行崩塌機率的預測，並將結果進行可視化流程，達到即時於畫面呈現可能崩塌範圍之成效(圖 4-7)。

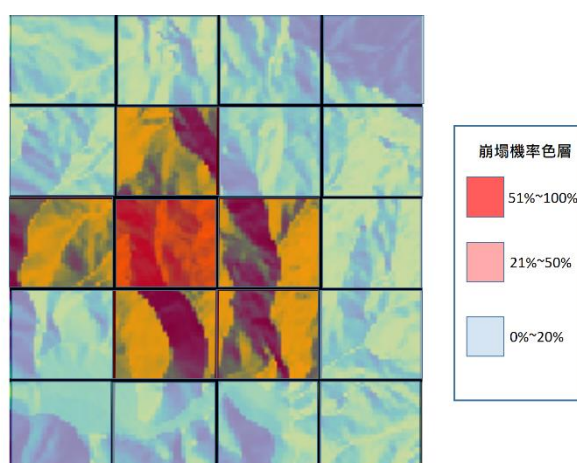


圖 4-7 崩塌機率預測結果可視化

資料來源：本計畫繪製

第五節 LSTM 模型建置與訓練結果

研究並了解 LSTM 的建置模式後，使用資料庫中的地震時間序列以及降雨時間序列資料，以歷史崩塌事件以及其範圍做為訓練目標，透過增加樣本數以及數值處理等方式，訓練 LSTM 模型，找出最佳化模型後對時間序列進行因子解算。解算出時間序列因子後，將其結合地質特性、即時地表情況等環境因子進行崩塌機率的預測。建立前饋式類神經網路模型，使用資料庫中崩塌發生之歷史事件為訓練資料，隨機取樣崩塌歷史事件的 80%做為訓練資料，其餘 20%做為測試資料。訓練一類神經網路模型之權重，使其能準確結合各項因子來預測崩塌機率。使用的因子包括地文因子以及由長短期記憶模型所得之時間序列因子，進行坡地發生崩塌之機率預測並顯示於網格資料中。

如果只依照一般方式收集降雨和地震震度資料，也就是把降雨資料收集日期範圍的 9 天分成三段 72 小時以及把地震資料收集日期範圍的 5 年分成三段至四段 72 周，會導致訓練資料數目完全不夠，所以決定採用期中報告中提出的概念將時間以 1 小時回推的方式創造出更多時間序列，降雨時間序列從原本的 3 筆 72 小時的資料擴大成 145 筆 72 小時的資料，將 9 天 216 小時先扣除最後 3 天 72 小時，剩下 144 小時，每 1 小時往後加 71 小時當作一筆資料，所以會有 144 筆 72 小時的資料數目，最後再加上一開始扣除的 72 小時那筆資料，總共會有 145 筆資料，也就是 $(9 - 3) \times 24 + 1 = 145$ ，而地震震度資料時間序列資料也依循相同的概念，將時間以 4 周回推的方式將資料數量擴大至 48 筆，將 5 年 260 周先扣除最後 72 周，剩下 188 周，第 1 周往後加 71 周當作一筆資料，再來是第五周往後加 71 周當作一筆資料，以此類推，會有 47 筆 72 周的資料數目，最後再加上一開始扣除

的 72 周那筆資料，總共會有 48 筆資料，也就是 $(5 \times 52 - 72) / 4 + 1 = 48$ ，為資料增加多樣性以提高模型的準確性。

除了增加訓練資料的多樣性，訂定有效的因子值也是可以提高模型準確度的方式，以下兩張圖分別是訂定降雨及地震震度時間序列資料因子值的方法(如圖 4-8、圖 4-9)，降雨時間序列資料是以 1 小時回推的方式收集資料，資料相對於地震資料精度較高，所以在因子值設定方法也較複雜。我們是以已發生的崩塌事件為資料預測未來崩塌事件機率，所以我們知道崩塌事件中崩塌網格及未崩塌網格各是哪些，一開始先將網格區分成崩塌網格及為崩塌網格，兩者所訂定的因子值會不同，在崩塌期間所設定的因子值分別為 1 和 0.5，設定崩塌網格在崩塌期間的因子值為 1 是合理的，而未崩塌網格設定為 0.5，崩塌期間的崩塌網格附近的網格都有一定的機率崩塌，再來將時間分為崩塌前 24 小時、48 小時及 72 小時，並分別設定不同的因子值數值，計算因子值時，需先知道網格崩塌與否，再觀察 72 小時的時間段落落在哪個區間，再以小時為單位計算，以崩塌網格舉例，當 12 小時落在崩塌前 48 至 72 小時之間，24 小時落在崩塌前 24 至 48 小時之間，24 小時落在崩塌日至崩塌前 24 小時，以及 12 小時落在崩塌期間，最後的計算結果為 $0.7(12*0.4+24*0.6+24*0.8+12*1) / 72$ 。地震震度時間序列資料因為資料精度較低，所以再設定因子值時的方法也較為簡單些，與降雨時間序列相同，一開始先將網格區分成崩塌網格及為崩塌網格，兩者所訂定的因子值會不同，再將時間分為崩塌前 52 周、156 周以及 260 周，三個時間區間崩塌網格所對應的因子值分別為 1、0.8 和 0.5，而為崩塌網格分別對應的因子值為 0.5、0.3、0，計算因子值的概念及步驟與降雨時間序列相同。

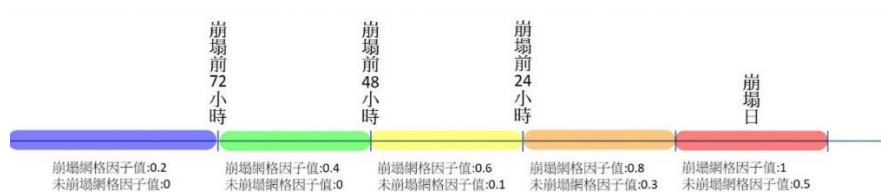


圖 4-8 降雨因子值設定方法

資料來源：本計畫繪製

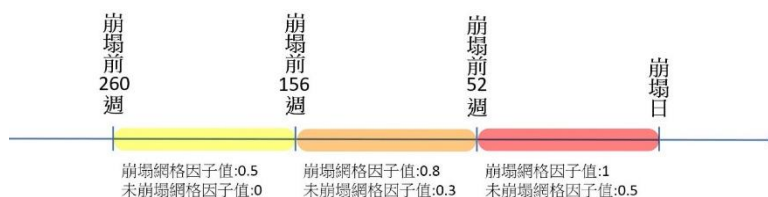


圖 4-9 地震震度因子值設定方法

資料來源：本計畫繪製

目前訓練的 1273 筆網格資料來自花蓮、台中、南投、新北以及嘉義，其中包含 606 格崩塌網格及 667 格未崩塌資料，依上述方法得到加權後的雨量資料，每個地區都各有 145 筆 72 小時的降雨時間序列資料，所以 5 個地區總共會有 725 筆資料。地震資料而是每個地區 49 筆 72 周的地震震度時間序列資料，所以總共會有 245 筆資料。

上述之每筆資料都包含對應地區之網格數，以花蓮為例，1 筆降雨時間序列資料包含 144 格網格 72 小時的加權時降雨量，1 筆地震震度時間序列列資料包含 144 格網格 72 週的地震震度資料。降雨及地震震度 LSTM 訓練目前都是以每個地區 10 筆資料，總共 50 筆資料下去訓練，網格資料總數為 12730 組，訓練過程資料會被隨機分成 10184 筆訓練資料及 2546 測試資料。

透過訓練得出第一次降雨訓練圖(如圖 4-10) 與地震震度訓練圖(如圖 4-11)，藍色及橘色線分別代表訓練集和驗證集，在理想情況下

我們會希望兩條線能收斂至同一 Loss 值(也就是 Mean Square error)，將 epoch 設定為 30，從降雨訓練圖中可以看出訓練過程中訓練集和驗證集多次上下震盪，但最後 Loss 值都收斂到 0.036 左右，而地震震度訓練圖的訓練集和驗證集在訓練過程中也多次上下震盪，但震盪幅度沒有降雨訓練圖大，最後也沒有明顯收斂的狀況，以結果來說不算

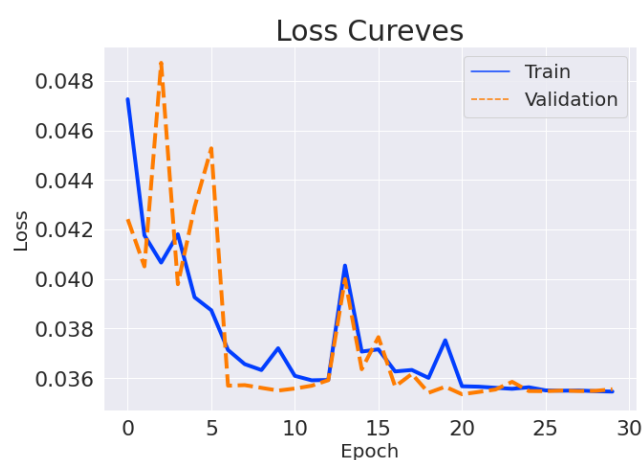


圖 4-10 第一次降雨訓練圖

資料來源：本計畫繪製

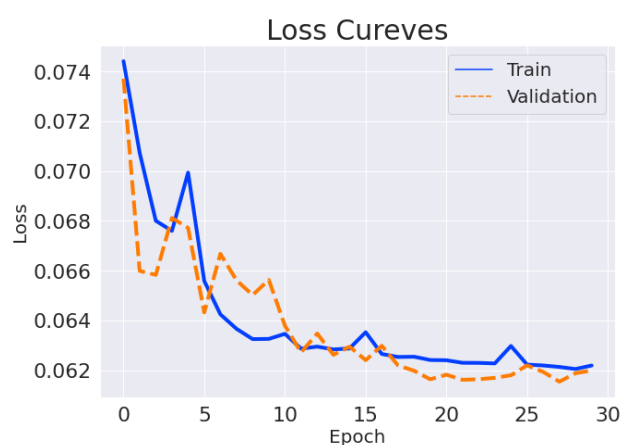


圖 4-11 第一次地震震度訓練圖

資料來源：本計畫繪製

重新訓練得出第二次降雨訓練圖(如圖 4-12) 與地震震度訓練圖(如圖 4-13)，epoch 設定為 30，從降雨訓練圖中可以看出訓練過程中訓練集和驗證集多次上下震盪，但結果與第一次訓練不同，最後並沒有呈現收斂的狀況，地震震度訓練圖的訓練集和驗證集在訓練過程中也多次上下震盪，最後訓練集有稍微呈現收斂但驗證集還在震盪，以結果來說並不是一個好的訓練結果。

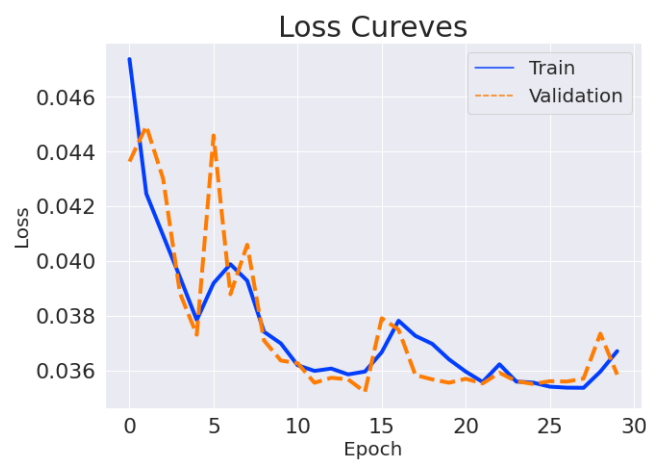


圖 4-12 第二次降雨訓練圖

資料來源：本計畫繪製

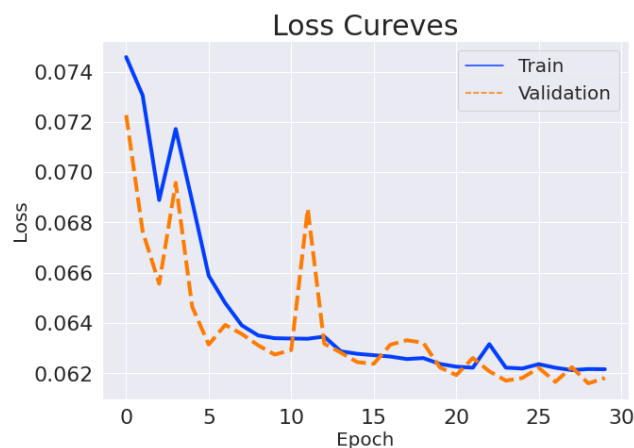


圖 4-13 第二次地震震度訓練圖

資料來源：本計畫繪製

重新訓練得出第三次降雨訓練圖(如圖 4-14) 與地震震度訓練圖(如圖 4-15)，epoch 設定為 30，從降雨訓練圖中可以看出訓練過程中訓練集和驗證集多次同時上下震盪，結果與第一次訓練類似，最後呈現微幅震盪至收斂情況，地震震度訓練圖的訓練集和驗證集在訓練過程中多次上下大幅震盪，而且最後兩者都沒有降至前兩次訓練的最終 Loss 值 0.062 左右，以結果來說降雨訓練算是一個不錯的結果，但還可以再精進，地震震度訓練是一個很差的訓練結果。

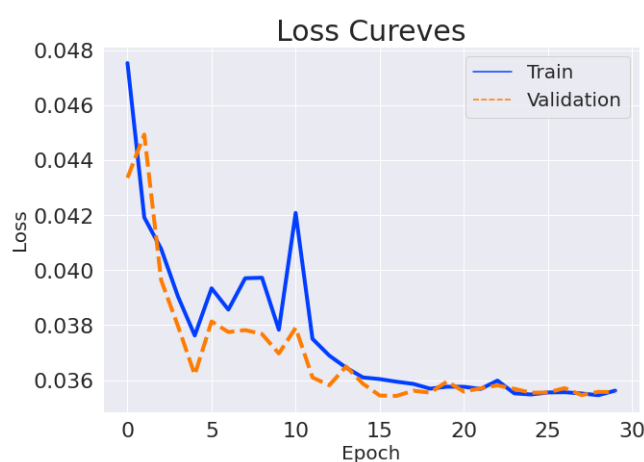


圖 4-14 第三次降雨訓練圖

資料來源：本計畫繪製

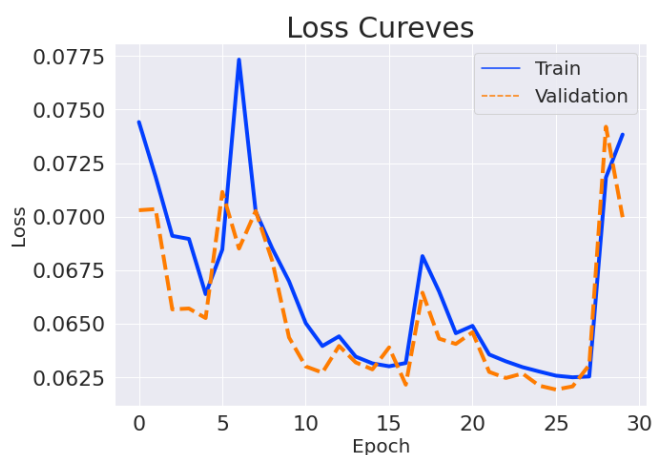


圖 4-15 第三次地震震度訓練圖

資料來源：本計畫繪製

觀察前三次訓練的結果，都沒有發現好結果，所以將第四次的訓練的 epoch 調高至 500 觀察訓練成效，訓練得出第四次降雨訓練圖(如圖 4-16) 與地震震度訓練圖 (如圖 4-17)，從降雨訓練圖中可以看出訓練過程中訓練集和驗證集多次上下震盪，而且可以發現兩者呈現下降趨勢，從訓練集和驗證集的最終 Loss 值可以更清楚的看出相較前三次於前三次訓練低，雖然最後兩者並沒有收斂至相同 Loss 值，但這個方向是可行的，地震震度訓練圖的訓練集和驗證集在訓練過程中多次上下小幅震盪，兩者也呈現下降趨勢，且最終 Loss 值相較於前三次訓練低，最後訓練集和驗證集趨近收斂至一值，這些現象顯示訓練方向正朝著好的方向前進，以結果來說降雨訓練是一個不錯的結果，但還可以再精進，地震震度訓練是一個好的訓練結果。

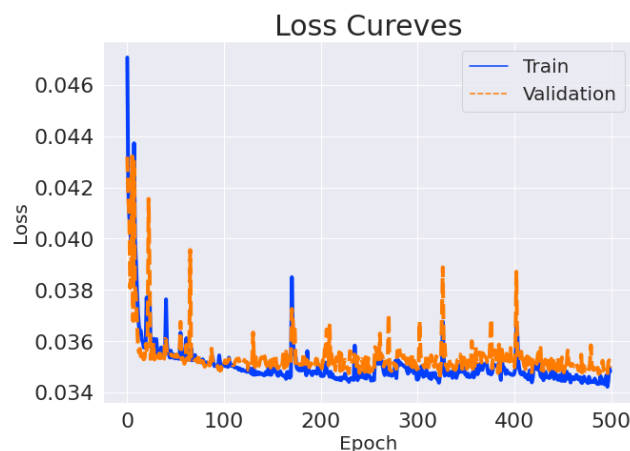


圖 4-16 第四次降雨訓練圖

資料來源：本計畫繪製

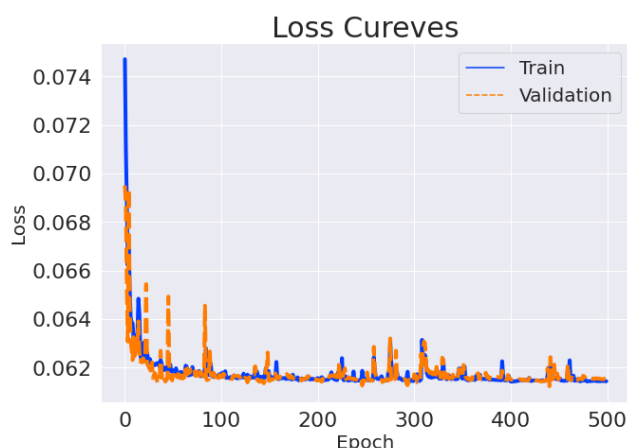


圖 4-17 第四次地震震度訓練圖

資料來源：本計畫繪製

第五次訓練照著增加 epoch 的方向繼續努力，將 epoch 增加到 1000，訓練得出第五次降雨訓練圖(如圖 4-18) 與地震震度訓練圖 (如圖 4-19)，觀察第五次降雨訓練圖後，發現朝著增加 epoch 的方向是有效果的，除了維持震盪下降的趨勢外，最終 Loss 值也降至 0.034 左右，但也發現驗證集的下降趨勢沒有訓練集那麼明顯，導致兩者最終 Loss 值無法收斂至一值，在第四次訓練也有發現相同問題，地震震度訓練圖的結果和第四次訓練成果相似，呈現一個震盪下降的趨勢，訓練集與驗證集的下降趨勢大致相同，且最終 Loss 值都收斂至 0.061 附近，經過第四次與第五次訓練，增加 epoch 會是未來將 Loss 降低的方向，會增加至訓練圖出現過度訓練的趨勢，也就是訓練集持續呈現下降趨勢，而測試集停止下降，開始出現上升趨勢，第五次訓練中降雨訓練圖就有出現這個問題，但還不明顯，未來會嘗試至出現較為明顯的過度訓練，地震震度目前還沒出現過度訓練，所以 epoch 可以再繼續增加，目前是使用第五次訓練的模型當成最終預測因子值的模型，觀察降雨 LSTM 訓練集、測試集(如圖 4-20、圖 4-21)以及地震震度

LSTM 訓練集、測試集(如圖 4-22、圖 4-23)可以發現大部分的點分布沒有落在斜率為一的紅色需線，當訓練結果越好，點的分布會越靠近紅色線。

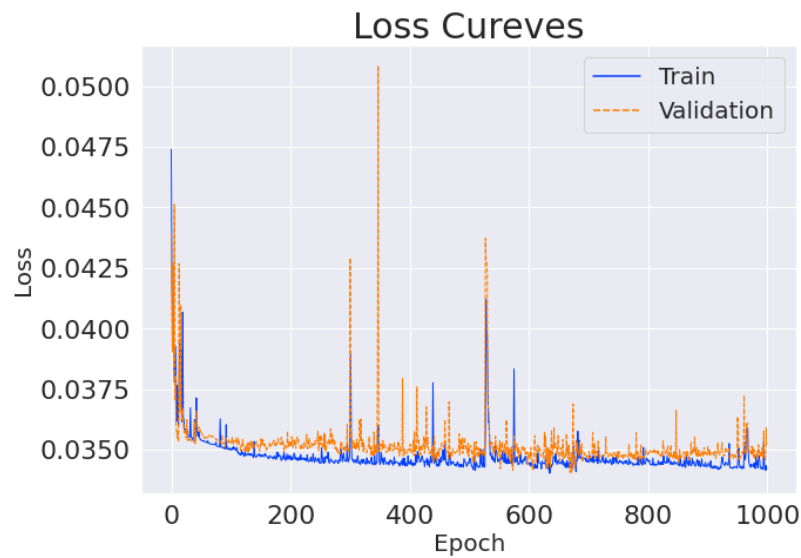


圖 4-18 第五次降雨之訓練圖

資料來源：本計畫繪製

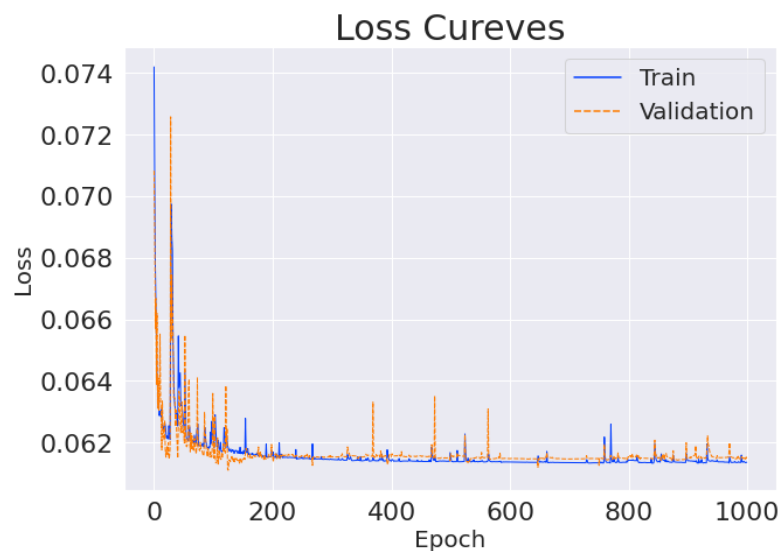


圖 4-19 第五次地震震度訓練圖

資料來源：本計畫繪製

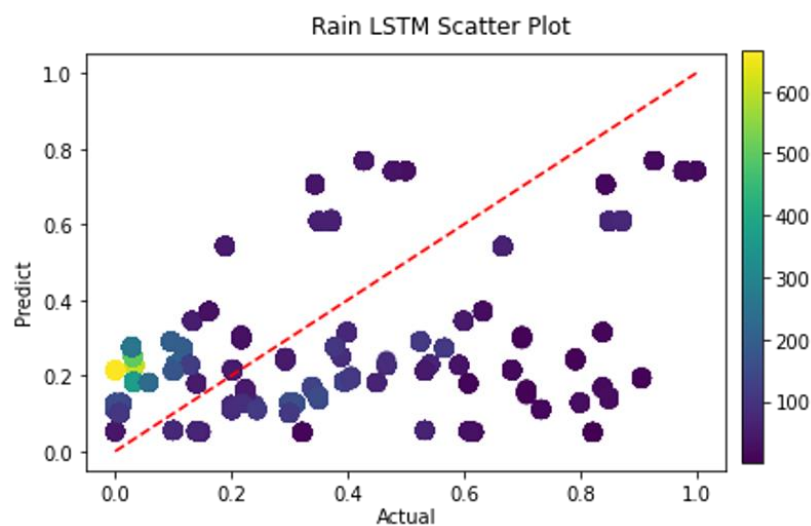


圖 4-20 降雨 LSTM 訓練集

資料來源：本計畫繪製

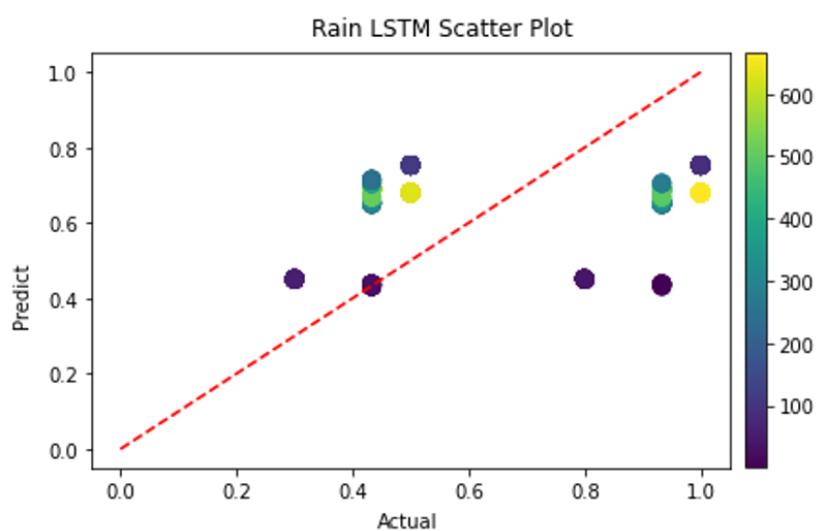


圖 4-21 降雨 LSTM 測試集

資料來源：本計畫繪製

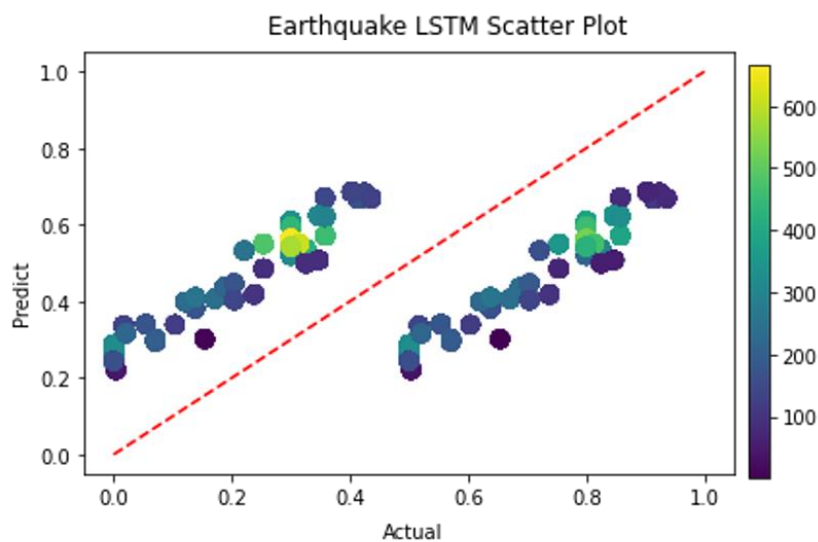


圖 4-22 地震震度 LSTM 訓練集

資料來源：本計畫繪製

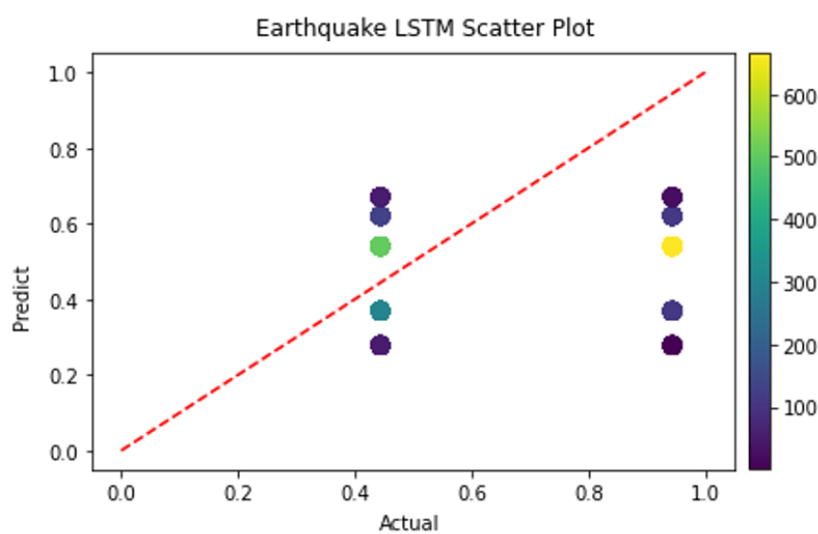


圖 4-23 地震震度 LSTM 測試集

資料來源：本計畫繪製

第六節 ANN 模型建置與訓練結果

將訓練完成的降雨時間序列 LSTM 模型匯出，並以此預測 ANN

中 1273 格網格的降雨因子值，原本是以崩塌結束日往前 216 小時中之連續 72 小時作為降雨 LSTM 的輸入資料，改成以崩塌結束日前 72 小時做為輸入資料，即可預測出的崩塌結束日前 72 小時的降雨因子值(如表 4-1)。

將訓練完成的地震震度時間序列 LSTM 模型匯出，並以此預測 ANN 中 1273 格網格的地震震度因子值，原本是以崩塌結束日往前 260 週中之連續 72 週作為地震震度 LSTM 的輸入資料，改成以崩塌結束日前 72 週做為輸入資料，即可預測出的崩塌結束日前 72 週的地震震度因子值(如表 4-1)。

表 4-1 降雨與地震震度預測因子值

	降雨預測因子值	地震震度預測因子值
1	0.753069699	0.620396256
2	0.753069699	0.620396256
3	0.75306952	0.620396256
4	0.75306952	0.620396256
5	0.753069341	0.620396256
6	0.753069341	0.620396256
7	0.753069222	0.620396256
8	0.753069162	0.620396256
9	0.753069162	0.620396256
10	0.753069043	0.620396256

資料來源：本計畫整理

ANN 的環境因子就如第三章所述，坡向是利用 QGIS 軟體內建之程式分析，使用周邊 8 格網格推斷出中間網格的坡向數據，最後得出每格網格對應之坡向數值，坡度也是利用 QGIS 軟體內建之程式分析，根據範圍內的坡度分級和面積占比之乘積決定該網格範圍之平均坡度，最後得出每格網格對應之坡度數值，高程是利用 GRD 檔進行

檔案轉換後得出之網格高程數值，而植被狀況是利用 EO Browser 中的 Sentinel-2 衛星中所記載的 NDVI 網格資料，以及告訴程式此網格崩塌與否的崩塌數據，崩塌網格紀錄為 1，未崩塌網格紀錄為 0，最後配合上降雨和地震震度因子值即為 ANN 訓練資料(如表 4-2)。

表 4-2 ANN 訓練資料

	降雨因子值	地震震度因子值	坡度	坡向	高程	NDVI	崩塌
1	0.753069699	0.620396256	40.88310242	338.9377747	2072.03	0.531618465	0
2	0.753069699	0.620396256	19.81529617	355.1035156	2056.92	0.66925734	0
3	0.75306952	0.620396256	34.76631927	143.5009918	2054.91	0.68444449	0
4	0.75306952	0.620396256	49.58966827	143.1634522	2057.12	0.69237256	0
5	0.753069341	0.620396256	48.24071884	133.4126282	2054.53	0.62590286	0
6	0.753069341	0.620396256	48.58384323	130.0665741	2065.42	0.63637784	0
7	0.753069222	0.620396256	48.45828247	126.325119	2078.19	0.69573358	0
8	0.753069162	0.620396256	49.19252777	124.592926	2099.91	0.68787384	0
9	0.753069162	0.620396256	53.09357071	130.9311066	2116.83	0.627608003	0
10	0.753069043	0.620396256	56.63201141	135.153244	2137.57	0.685308755	0

資料來源：本計畫整理

ANN 訓練過程會將 1273 比數據分成 1018 筆訓練資料和 255 筆驗證資料去建立模型，epoch 設定為 1000，將訓練完成的模型匯出，以相同的 ANN 訓練資料作為輸入資料去預測崩塌機率，最後得出的 1273 格網格預測崩塌機率(如表 4-3)。

表 4-3 網格預測崩塌機率

	預測崩塌機率
1	0.168041289
2	0.009230353
3	0.432273269
4	0.282905191
5	0.342261642
6	0.338175625
7	0.312283695
8	0.310419023
9	0.26262489
10	0.168041289

資料來源：本計畫整理

透過不同臨界點的混淆矩陣圖(如圖 4-24)即可觀察出設定不同臨界點時真陽性、偽陽性、偽陰性、真陰性的分配情況，透過觀察接收者操作特徵曲線(如圖 4-25)也就是 ROC 曲線也可以判斷此模型的預測能力，當 auc 數值越高顯示此模型正確率越高。

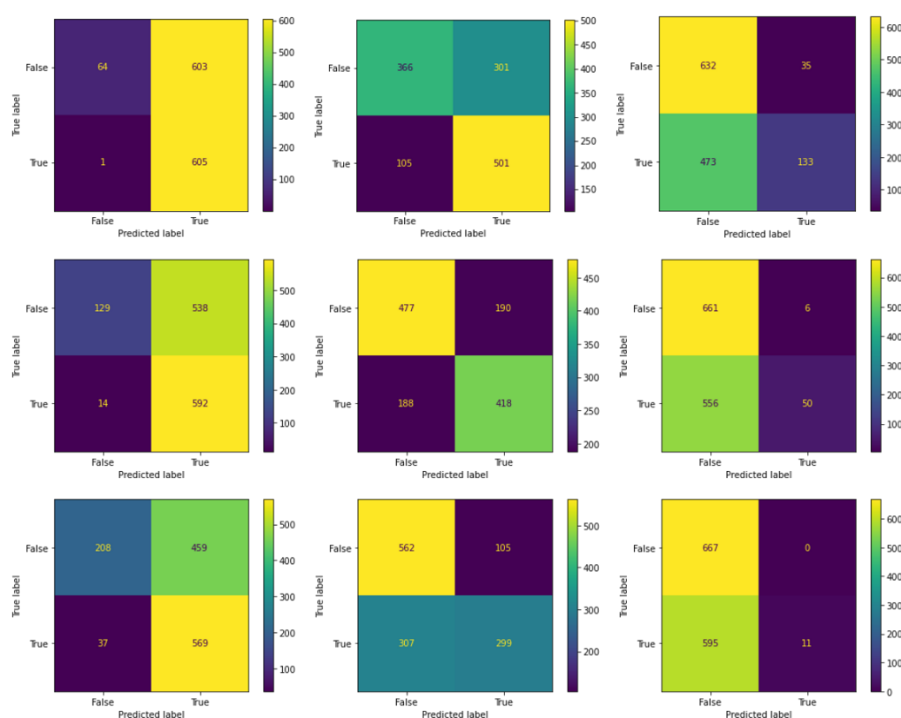


圖 4-24 混淆矩陣圖

資料來源：本計畫繪製

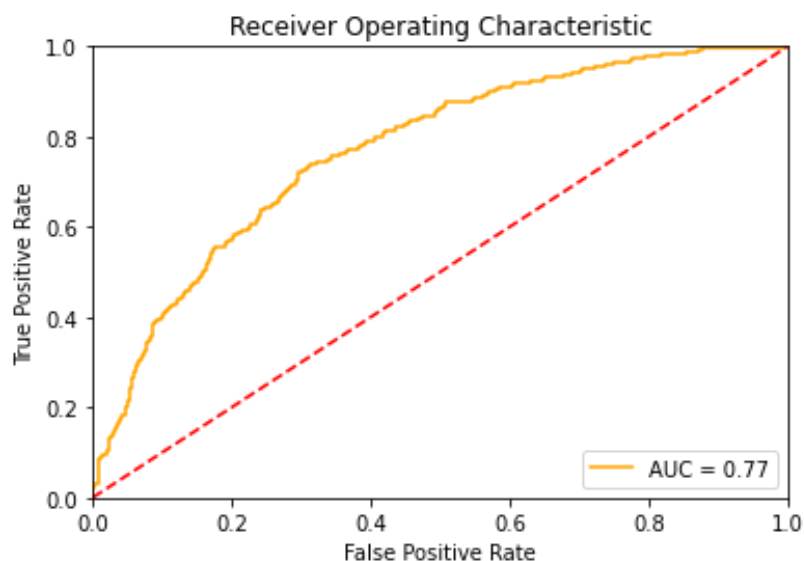


圖 4-25 ROC 曲線圖

資料來源：本計畫繪製

此模型的 auc 值為 0.77，當 auc 值大於 0.5 時顯示模型有預測能力。設定不同臨界點之準確率、敏感度、特異度數據如下表 4-4 所示。

表 4-4 混淆矩陣數據

臨界點	準確率	敏感度	特異度
0.1	0.526	0.998	0.096
0.2	0.566	0.977	0.193
0.3	0.610	0.939	0.312
0.4	0.681	0.827	0.549
0.5	0.703	0.690	0.715
0.6	0.676	0.493	0.843
0.7	0.600	0.219	0.948
0.8	0.559	0.083	0.991
0.9	0.533	0.018	1

資料來源：本計畫整理

利用本計畫建置之模型，輸入任一網格任一時間點前之地震及降

雨歷時，以及該網格之地形及地貌特性，即可預測該網格崩塌機率，重複進行後則可獲得區域性崩塌潛勢預測，如圖 4-26 所示，其中顏色越淺之網格表示有越大崩塌可能性，預測結果與實際崩塌事件相比具有相似崩塌分佈趨勢。

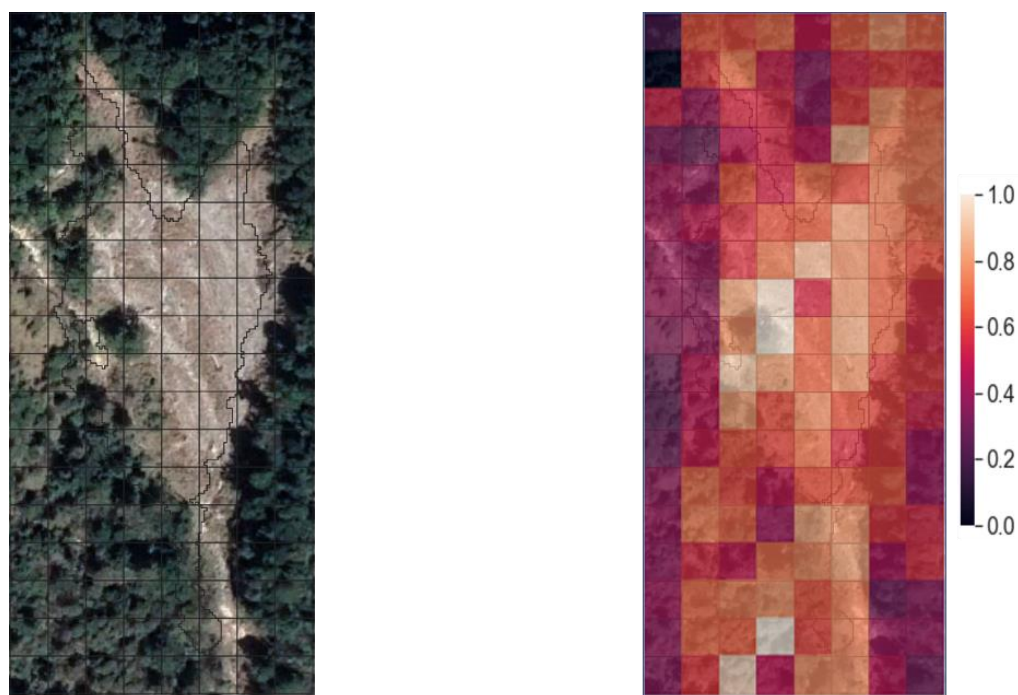


圖 4-26 崩塌預測圖

資料來源：本計畫繪製

第七節 小結

目前所知增加 ANN 模型預測能力的方法如下，第一點為訓練 LSTM 模型時將 epoch 提高，以每次 500 為單位增加直到模型出現過度訓練的趨勢，從第四節得知增加 epoch 可以降低模型的 Loss 值，目標將 Loss 值降至 0.03 以下。

第二點為增加 LSTM 模型訓練次數，當 epoch 提高到模型出現過度訓練的趨勢時即將 epoch 固定，開始增加 LSTM 模型訓練次數，當

訓練出的模型達到一定數量，即可從中挑選出預測能力較佳的 LSTM 模型，並用以預測降雨和地震震度因子值。

第三點為增加網格數量，計畫最後預計交付台灣全地區崩塌預警系統，目前只有 5 個地區總共 1273 格網格，且缺乏南部崩塌網格資料，目前崩塌網格及未崩塌網格的樣本數並不足夠製作出台灣全地區崩塌預警系統，當樣本數提高後即可降低各地區網格資料對總體網格資料的影響。

第四點為增加網格環境因子多樣性，土壤質地、順向逆向波等影響崩塌之因素都應納入環境因子的收集範圍，多樣的環境因子幫助模型在訓練時在參數的選擇性增加，達到更平衡的權重配置，減少誤差。

第五章 結論與未來展望

第一節 結論

由於台灣因地理位置地震發生頻繁而易於產生敏感地層帶，配合著梅雨及颱風強降雨，造成的崩塌地質災害數量不在少數，因此本計畫提出以人工智慧技術建立一崩塌機率預測模型，選用台灣崩塌發生頻繁的南投、台中、花蓮等地區作為對象，首先進行一系列網格資料收集以及 LSTM 和 ANN 模型建置與測試流程，透過資料的量化與分析結合人工智慧相關技術建立一有效之預測模型，提供各區域崩塌發生機率預測之有效依據。

本計畫已完成預測模型之建立，將時降雨量歷時及地震震度歷時以 LSTM 解算網格時間序列因子值，配合坡度、坡度、高程等網格環境因子資料，輸入後即會產生一崩塌預測值，目前準確率最高來到 73.8%。將計畫中兩 ANN 模型進行比較發現使用誤差較高的 LSTM 所建置之模型反而在準確率上優於使用誤差較低的 LSTM 所建置之模型。

目前研究結果顯示，以 LSTM 解算時間序列因子值、以 ANN 預測網格崩塌機率，皆能順利進行，惟現階段建立之模型尚有準確率不足的問題，若未來持續收集資料、訓練模型，將有助於提高預測模型的準確率。

第二節 未來展望

崩塌機率預測網路訓練完成後，能夠依據降雨資料和地震歷史所構成之時序因子以及由地質特性、地表情況所構成之環境因子，對每格網格資料進行崩塌機率預測，提供可能的受災範圍與規模，達到即時崩塌預警與救災的目的，將更進一步保證人民的生命財產安全，有效降低強降雨發生時之潛在地質災害所帶來的危害。利用前饋式神經網路探討不同降雨時序和地震時序對不同坡地條件上的影響，並預測降雨過後崩塌災害發生機率，將有利於未來強降雨事件或是地震事件發生後調整坡地災害警戒發布值，並提供潛在崩塌資訊，作為災害預警及即時救災之依據。因此後續規劃進行之工作項目如下：

- 一、持續建立資料庫與訓練臺灣北中南東不同區域之人工智慧模型建置，進行即時預測以網格為單位之崩塌發生機率，並即時顯示在介面上供決策者參考應用。
- 二、根據不同區域之潛勢分析結果，調整警戒值或防災觸發門檻，建立因地制宜的防災規劃。
- 三、將 LSTM 訓練時的 epoch 提高，達到降低 Loss 值之效果。
- 四、增加 LSTM 模型訓練次數，增加訓練模型數量即可從中選擇預測能力較佳之模型，並用以預測降雨和地震震度因子值

五、增加網格數量，當樣本數提高後即可降低各地區網格資料對總體

網格資料的影響

六、增加網格環境因子多樣性，未來計畫加入土壤質地、順向逆向波

等引響崩塌之因素。

參考文獻

- [1] Gers, Felix. Long short-term memory in recurrent neural networks. Diss. Verlag nicht ermittelbar, 2001.
- [2] Graves, Alex. "Generating sequences with recurrent neural networks." arXiv preprint arXiv:1308.0850 , 2013.
- [3] Hebb D. O. "The organization of behavior: A neuropsychological theory", John Wiley & Sons, New York, 1949.
- [4] Kawakami, Kazuya. "Supervised sequence labelling with recurrent neural networks." Ph. D. thesis, 2008.
- [5] Rosenblatt F. "The perceptron: A probabilistic model for information storage and organization in the brain", Psychological Review, Vol. 65, No. 6, pp. 386-408, 1958.
- [6] 苟程成、秦宇君、田甜、伍大勇、劉悅、程學旗(2017)。「一種基於RNN的社交消息爆發預測模型」。軟體學報，28，頁 3030-3042。
- [7] 張斐章、黃浩倫、張麗秋(2001)。「回饋式類神經網路於河川流量推估之應用」。農業工程學報，47，頁 32-39。
- [8] 鄭允中(民 106)。「基於長短期記憶遞迴類神經網路之新台幣兌美元匯率預測模型」。國立台灣大學，碩士論文。
- [9] Hochreiter S., & Schmidhuber J. "Long short-term memory", Neural Computation, Vol. 9, No. 8, pp.1735-1780, 1997.
- [10] 小林村事件簿。檢自
<https://sites.google.com/site/xiaolincunshijianbu/can-jie-lun/er-xiao-lin-cun-zai-bian-fa-sheng-shi-de-qing-kuang>(Apr.10,2010)
- [11] 交通部中央氣象局。檢自
<https://unfccc.saveoursky.org.tw/nir/2018nc/uploadsnc/filesfigure/5>
- [12] 石岡壩集水區保育實施計畫，計畫內容(民 110 年 3 月)。檢自
[https://www.archives.gov.tw/UserFiles/file/%E5%8F%83%E8%80%83%E6%96%87%E7%8D%BB%E6%A0%BC%E5%BC%8F\(105%](https://www.archives.gov.tw/UserFiles/file/%E5%8F%83%E8%80%83%E6%96%87%E7%8D%BB%E6%A0%BC%E5%BC%8F(105%)

E9%81%A9%E7%94%A8).pdf

- [13] 觀測資料查詢系統。檢自 <https://e-service.cwb.gov.tw/HistoryDataQuery/>
- [14] 內政部數值地形模型加值應用服務。檢自 <https://dtm.moi.gov.tw/aspect.aspx>
- [15] 台灣活動斷層。檢自 <https://fault.moeacgs.gov.tw/Gis/Home/pageMap>
- [16] 地質雲加值應用平台。檢自 <https://www.geologycloud.tw/map/Stratum/zh-tw>

附錄

附錄一、期初審查會議紀錄暨回覆辦理情形

項次	審查意見	回覆辦理情形
報告內容審查意見：		
壹、張竝瑜委員		
一	使用地震紀錄作 LSTM 時間序列輸入資料，須注意地震震源深度、破裂方式(應力)及場址效應均有不同，主持人要將距離多少之地震作為輸入訓練資料須加以討論與測試。加以震央距離非震源距離，且地震波傳遞途徑亦不同，因此考量應加以特別注意。	對於地震歷史紀錄之使用，本研究目前採取距網格中心最近之地震測站資料，資料的篩選使用距網格中心 30 公里內且規模大於 3 之地震做為考量之時間序列。
二	加入地震強度加速度作為輸入 FNN 需量化方能進行計算。因此，如何將岩性及岩層方向(走向、傾角)等環境因子量化，也應加以考慮。	研究目前已利用 DEM 計算量化高程、坡向及坡度等地形特性指標，然而這次報告中尚無使用岩石特性之資料，未來將視訓練狀況逐步將岩石特性納入資料庫內，岩層之走向及傾角規劃以類似地形特性計算量化指標；而岩性方面則會統整常見種類以在模型中加入分類變量進行分析。
貳、楊明德委員		
一	在 LSTM 模型中，地震與降雨的 time step 有較大差異，如何彙集在輸入資料。	本計畫針對地震及降雨分別訓練兩個 LSTM 模型，利用個別模型預測出之地震及降雨因子再整合於用於崩塌機率預測之類神經網路中，如此即可整合不同時間步長之歷時資訊。
二	深度學習模型之輸出內容為機率而非分類，因其為連續資料格式，所以需求更大量訓練資料，是否可克服此困難？	本計畫目前使用 100 年到 110 年的事件型崩塌目錄中隨機選取崩塌歷史紀錄，包含崩塌地及其周圍未崩塌區域，並以

項次	審查意見	回覆辦理情形
		網格為單位，目前包含崩塌網格 1367 個及未崩塌網格 2426 個，每一網格為一筆資料，以目前數據量而言訓練模型算是足夠，惟為考慮足夠資料多樣性及提升模型預測力，計畫將持續擴展現有資料集。
參、葉昭憲委員		
一	P.1：提列坡地崩塌發生特性；但未將「土地利用」納入模式輸入因子之原因為何？	謝謝委員的建議，本研究認為土地利用其內在自然特徵對坡地災害之影響大於人類對土地利用改造之形式，因此針對土地利用將使用植被狀況進行分析。植被狀況之數值量化常見方法一為計算區域內的植被覆蓋，另一種類為植生指標，計畫將採用 NDVI(常態化差異植生指標)、EVI(增強植被指標)，或 RVI(紅光比植生指標)。
二	P.4：圖 1.3-1 執行流程顯示輸出值為「潛在風險」然歷史資料崩塌結果為 0 或 1 之二元數值，此方法可自動轉換為機率值？	計畫規劃進行二分分類法方式訓練模型，以崩塌為 1、未崩塌為 0 方式訓練模型，模型進行預測時會給出借於 0 到 1 之分數，一般二分分類法在此步驟將設定閾值，大於閾值之預測為 1 反之為 0。此分數亦反應模型對崩塌機率之判斷因此可在設定閾值前使用此分數作為破壞機率基準
三	P.10：已選定研究範圍，而所使用資料年限為何？	本計畫目前使用 100 年到 110 年的事件型崩塌目錄中蒐集崩塌歷史紀錄，包含崩塌地及其周圍未崩塌區域，並以網格為單位，目前包含崩塌網格 1367 個及未崩塌網格 2426 個。並針對各網格蒐集其崩塌

項次	審查意見	回覆辦理情形
		前 72 小時之時雨量資料及近十年來規模為 3 以上且距離目標測站 30 公里以內之地震資料。
四	P.12：風險評估可否以豪雨、豪大雨等既定降雨條件予以呈現？	可以，訓練完成模型在給定降雨、地震因子及環境因子下將可預測崩塌機率，在已知降雨等級條件下可使用該等級降雨強度期望值計算降雨因子作為模型輸入計算崩塌風險。
肆、劉正千委員		
一	準確的坡地崩塌風險評估，是水土保持局防災工作重要資訊。若能引進人工智慧技術，善加分析已累積之巨量資料，提升風險評估的準確性，裨益甚大，值得深入研究。	謝謝委員支持，隨機器學習及大數據蒐集儲存技術突飛猛進，人工智慧技術已廣泛應用於各領域並證實可快速估算結果並維持可靠準確性，這也是本計畫採取此方法之初衷。
二	應用人工智慧深度學習架構去解決各種問題的成功關鍵，是必須要有數量龐大，而且具有包含了各種情況的資料庫來加以訓練，目前只選定石岡壩集水區作為本研究之範圍，事件的樣本足夠嗎？空間範圍有包含各種情況嗎？時間範圍足以驗證長短期的記憶嗎？	期初報告書中原本採用台中石岡壩附近地區做為崩塌資料收集對象，然而在資料過程中發現，100 年到 110 年的事件型崩塌目錄中，台中石岡壩附近的崩塌資料不足，本計畫目標是建立一崩塌預警系統，而使用範圍為台灣地區，所以在資料庫建立上，多樣性會是一個要點，在 QGIS 中觀察這 10 年間的崩塌資料後，決定選擇以台中、屏東、花蓮、新北、南投及嘉義為建置崩塌網格資料庫的地區。而針對時間範圍而言，計畫初步蒐集各網格前 72 小時之時雨量資料及近十年來規模為 3 以上且距離目標測站 30 公里以內之地震資料，如 72 小時降雨資料模型表現欠佳則會使用更長

項次	審查意見	回覆辦理情形
		期如六日降雨、雙周降雨或月降雨歷時。
三	本計畫執行需要搜集的很多資料，在水保局的 BigGIS 系統上都已經彙整，可以申請下載，建議不要重複投注人力、時間與資源去蒐集，待盤點現有資料後，確定是過去沒有的資料，再挹注資源去蒐集，也可以回饋給水保局納入 BigGIS，供其他人使用。	謝謝委員提供寶貴資訊，計畫會多加使用現有系統及資料庫並提供回饋。
四	目前選定的幾種環境因子中，土壤滲透率和植被狀況都是需要另外進行處理計算才能夠得到，而處理的方法也牽涉到許多可能導致誤差的來源，請問要如何定量評估所產製的這兩項環境因子？又會對人工智慧訓練的結果有何影響？請具體說明。	降雨及地震資料蒐集及各種環境因子之量化皆有隱含誤差而導致模型預測產生誤差之可能。一般而言可視其為隨機誤差，而誤差大小對模型表現之影響可使用誤差傳播理論來估算，並以預測區間顯示模型預測之不確定性。
伍、陳振宇委員		
一	時間序列因子中，地震建議採用PGV 或PGA 等震度資料，而非使用規模；另引致崩塌之雨量，過去研究常考慮3 小時及 24 小時雨量，亦可嘗試。	謝謝委員的建議，本研究將會以相同於規模之方式將 PGV 及 PGA 考入於模型輸入變量。雨量方面，目前採取 72 小時之降雨歷時，如較接近崩塌時間點之降雨資訊對崩塌有更直接影響力則人工智慧模型訓練過程中會提取該訊息以計算降雨因子，因此目前設定即考慮 3 小時及 24 小時雨量。
二	環境因子中，建議納入與斷層之距離。	謝謝委員的建議，本研究亦認知崩塌點與斷層距離為重要指標，會將其考慮於模型輸入，並藉由 LSTM 提取地震因子。

項次	審查意見	回覆辦理情形
三	學習及驗證均需大量崩塌實際案例，惟崩塌發生時間不易取得，這部分如何克服？另，水保局自 99 年起有提供事件型崩塌目錄之開放資料，或可參考。	謝謝委員提供之資訊。使用數據驅動方法之研究可能限於資料不足限制，解決方法之一為擴大現有資料集。使用水保局山崩目錄或甚至他國山崩資料，在掌握足夠環境變數情況下即可不受時空限制擴大資料集；另一方法是在現有資料集下尋找、強化環境變數之使用，使模型可更準確進行預測。
陸、技研小組(詹婉妤)		
一	時間序列因子建議增加土壤含水量進行坡地崩塌機率評估。	謝謝技研小組建議，本研究後續蒐集地質資料將會將土壤含水量納入考量。
二	有關預測崩塌發生率之成果會於何種介面呈現？	本計畫尚未規劃建立供公眾使用之平台以呈現崩塌發生率，目前將以呈現模型預測力為主，模型預測之崩塌發生率則會以網格型式顯示崩塌機率套疊於地圖上。

附錄二、期中審查會議紀錄暨回覆辦理情形

項次	審查意見	回覆辦理情形
報告內容審查意見：		
壹、楊明德委員		
一	本研究AI型之訓練集內容應先敘明，若只取用崩塌歷史事件做為訓練資料，所訓練之模型是否趨於false positive之判釋？	報告中包含預測模型之混淆矩陣圖，除可用於模型表現外亦可判釋 false positive 及 false negative 之發生比例，有助於模型下一步改善策略。
二	坡向雖以數值化方式呈現，惟其影響與順向坡及逆向坡有關，如何將其納入AI模型之計算，請說明。	感謝委員的建議，未來計畫將順向逆向坡等環境因子納入訓練模型中，範圍中包含順向坡之網格，可考慮使用二元分類預測，將其順向坡因子值設為 1，其餘為 0；另一作法可考慮使用岩層走向及地表坡向之夾角作為因子以預測崩塌潛勢。
三	目前結果發現崩塌網格及未崩塌網格鑑別度不高，源自於降雨分布源於相同雨量站之加權內插差所致。建議可考量以坡地單元代替網格，且考慮地形及地質因子於模型中。若僅放大網格，雨量變異可能有限。	本計畫將降雨及地震震度歷時資料以推移的方式增加資料樣本數，雖然沒有根本上解決網格間鑑別度不足之問題，但樣本數數量提高，因此增加了模型的敏感度，間接解決了問題。
四	後續工作規劃要根據不同區域之潛勢分析結果調整警戒值，宜參考目前水保局之警戒值訂定原則，以便符合現地情況。	感謝委員的建議，未來計畫將朝這個方向出發，比較本計畫模型之預測機率與水保局之警戒值以結合二者做即時預警。

項次	審查意見	回覆辦理情形
貳、吳玉華委員		
一	研究範圍原選鄰近車籠埔斷層之石岡壩集水區作為本計畫之研究範圍，後改為選取臺灣北、中、南、東各地區崩塌相對密集的區域作為本計畫研究範圍之理由，請於報告書中補充說明。	石岡壩集水區附近崩塌樣本數不足，以相同地區資料訓練模型結果將出現嚴重的地區侷限性，因此本計畫考慮不同地區具不同特性之崩塌事件以訓練一通用模型。
二	地震歷史紀錄，除考量30公里內且規模大於3之地震，未考量地震深度之理由，請說明。	考量到實際情況，本計畫目前改以地震震度為時間序列因子，此考量即包含地震深度之影響。
三	植被狀況之數值量化將採何種方式，植生覆蓋率或植生指標。	本計畫以常態化差值植生指標(NDVI)為量化植被狀況之指標。
四	LSTM網路訓練和驗證，樣本資料80%為訓練樣本，20%為下載時間：SWCB(111年4月1日版)第2頁/共3頁測試樣本，樣本數多少以上才妥適，若只以一地區(花蓮)資料為訓練樣本，是否恰當？	本計畫目標為建立一有效之預測模型，提供各區域崩塌發生機率預測之有效依據若只以一地區資料為訓練樣本，模型結果將出現地區侷限性，因此訓練資料蒐集全台各地具不同條件之崩塌事件，以訓練通用模型，並於測試集樣本展示其通用預測之成果。
五	其他：報告書錯別字或謬誤請修正。 1. P3-6倒數第二行(圖4.4-1)(圖3-6)。 2. P3-7圖3-7坡度資料一級坡(重複二次)。 3. P3-12倒數第二行，”坡向圖”應為”坡向圖”。 4. P3-13倒數第3行、P3-14第1行、倒數第3行、第7行，”分佈”應為”分布”。” 5. P3-14圖3-19，”波度圖”應	謝謝委員的提醒，已經正錯報告內容別字及謬誤。

項次	審查意見	回覆辦理情形
	為”坡度圖”。	
	6. 植被或植披，請釐清(P3-8第3行、P3-16倒數第5行)。	
參、陳振宇委員		
一	P3-2，環境因子使用之崩塌歷史事件，依報告書的內容來看，應為「事件型崩塌目錄」，而非土石流防災資訊網之「重大土砂災例最速報」等調查報告，請修正。	謝謝委員的提醒，已修正。
二	P3-11，地震資料建議使用「震度資料」(如PGA及PGV)，而非使用地震規模，震度才是直接造成崩塌的原因。	謝謝委員的建議，考量到實際情況，本計畫目前改以地震震度為時間序列因子。
三	P4-7，當使用「事件型崩塌目錄」時，須瞭解其詮釋資料之意義，避免誤用。例如，2017年某崩塌圖資詮釋資料為「0601豪雨」，並非指6月1日發生崩塌，而是在6月1日”開始”的豪雨事件期間發生崩塌。如參考「重大災例最速報」，2017年0601豪雨事件，多數災害係發生於6/2~6/4間。	感謝委員的建議，已將崩塌時資料修改成正確的時間，未來收集資料時將參考重大災例最速報支內容。
四	目前採20m×20m網格進行訓練，因尺度較小對於環境因子及時間序列因子而言，大概只有地形會有差異，故將網格放大似乎為可嘗試的方向，唯過份放大又會降低地形的影響，宜多測試。	感謝委員的建議，本計畫將降雨及地震震度歷時資料以推移的方式增加資料樣本數，將維持以20m×20m的網格範圍進行資料蒐集。
肆、技研小組(李哲宇)		
一	P3-2，第二段提到植被為實際探查所獲得，P3-16說明利用衛星影像進行植被評估，請問後續打算採取現地調查還是衛星影像評估？	本計畫以常態化差值植生指標(NDVI)為量化植被狀況之指標。

項次	審查意見	回覆辦理情形
二	本次僅將坡度、坡向、高層資料列入環境因子，建議仍需將其他因子納入，至少依說明提到植被狀況可利用衛星影像判釋技術者進行進一步分析，並嘗試說明除已列入因子外，尚未納入之其他因子是否有文獻或貴單位說明採納困難或在整體影響占比重要性。	感謝委員的提醒，受限於計畫時長，本計畫目前將坡度、坡向、高程、植被狀況納入環境因子，目前模型已具相當預測能力(>70%準確率)，未來將繼續增加環境因子、增加多樣性，已進一步提升模型對崩塌機率預測之效果。

附錄三、期末審查會議紀錄暨回覆辦理情形

項次	審查意見	回覆辦理情形
報告內容審查意見：		
一	建議持續精進，俾未來納入相關業務應用。	謝謝委員建議，團隊認為此模型目前在僅考慮地震及降雨歷時及地形地貌特性及可有不錯的預測效果，如逐步增加如地質、岩層走向、地下水位、飽和度等影響因子，以及增加訓練資料、提升現有地形圖資精度，可獲得更好之模型預測表現，因此這也會是將來持續研究之目標。
二	1-4 頁圖 1-3 中僅只有地貌特性(此應指植被覆蓋)，是否需增加地形特性(此指坡度、坡向、高程)?	謝謝委員指出報告圖表疏漏，計畫起初欲考慮地形、地貌及地形特性，惟在時間限制下目前只包含地形及以植被為主之地貌特性。目前已跟據模型現況對該圖進行修正。
三	請詳細說明4-7 頁如何將降雨序列3 筆擴大成145 筆，與地震序列4 周擴大49 筆。	感謝委員的建議，降雨資料擴充過程為將 9 天 216 小時先扣除最後 3 天 72 小時，剩下 144 小時，每 1 小時往後加 71 小時當作一筆資料，所以會有 144 筆 72 小時的資料數目，最後再加上一開始扣除的 72 小時那筆資料，總共會有 145 筆資料，也就是 $(9 - 3) \times 24 + 1 = 145$ ，而地震震度資時間序列資料也依循相同的邏輯將 5 年 260 周先扣除最後 72 周，剩下 188 周，第 1 周往後加 71 周當作一筆資料，再來是第五周往

項次	審查意見	回覆辦理情形
		後加 71 周當作一筆資料，以此類推，會有 47 筆 72 周的資料數目，最後再加上一開始扣除的 72 周那筆資料，總共會有 48 筆資料，也就是 $(5 \times 52 - 72) / 4 + 1 = 48$ 。
四	4-8 頁請解釋文中「一開始先將網格分成崩塌網格及為崩塌網格」為甚麼意思？	謝謝委員指出報告解釋不詳細之處，此處為訓練降雨及地震 LSTM 模型，必須先設定模型預測目標值，目標值介於 0 到 1 之間，數值越高表示越接近崩塌狀態。團隊將崩塌事件內崩塌範圍網格化，則事件內實際崩塌之網格即為我們希望 LSTM 模型可預測出較高崩塌因子之網格，反之，未崩塌之網格則希望模型預測相對較低之崩塌因子。故在蒐集每一崩塌事件後，需要先對各網格進行標註，後續再由該地震或降雨歷時距離崩塌時間點之遠近，定義不同之目標崩塌因子，用以訓練模型。
五	4-20、4-21 頁文字敘述是否整理成表格讓人較容易閱讀其意思。	謝謝委員建議，已將此處不同臨界值對應之敏感度及特異度整理為表格，請見表 4-4。