

機器學習運用在劇烈降雨事件診斷預警上之研究

林得恩¹、張明瑞²

¹空軍氣象聯隊第20任聯隊長、^{1、2}國立臺灣大學氣候天氣災害研究中心

摘要

我國軍高山陣地，面對天氣系統所造成降雨時空分布的不均，以及每每打破紀錄的總累積降雨量，坡地災害將是未來倍加嚴峻的考驗與挑戰。

本研究的執行，主要目的為透過科學量化的研判步驟，對歷史坡地災害事件進行診斷分析，藉此釐清當時環境的關鍵氣象參數特徵，進而延長坡地災害可能肇生的事先預警整備時間，有效支援強化防災應變指揮官進行災害警戒調度的決心下達。

利用氣候統計及機器學習等研究方法，針對臺灣地區劇烈降雨進行個案研究。利用機器學習建模，建置穩定可靠的崩塌潛勢預報模式，提出最佳崩塌潛勢前兆之AI方法論，並有效回饋至部隊防災預警的實際整備作業上。

關鍵字：高山陣地、坡地災害、劇烈降雨事件。

1. 前言

近年來，氣候變遷加劇下，劇烈降雨天氣的肇生發展及強度變化，日趨嚴重與頻仍。我國軍高山陣地，面對天氣系統所造成降雨時空分布的不均，以及每每打破紀錄的總累積降雨量，坡地災害將是未來倍加嚴峻的考驗與挑戰。如何預先瞭解並掌握坡地災害潛勢的環境風險，有效釐清山區降雨特徵與強度，並事前完善預警整備，透過科學的作業與驗證，將暴烈天災可能造成的損失與傷害降到最低，俾利確保我國軍戰力發揚。

肇致坡地災害的3個重要條件，分別為一定的地形坡度、坡度上要有足夠的堆積物，以及充沛的降雨量（李等，2011）；尤其是降雨行為，更是整個事件肇生的驅動者。然而，降雨只是結果，環境條件以及氣象量場在整個發展過程的前兆時刻，行為特徵早已起了變化。從歷史氣候統計資料的彙整分析，可以有效瞭解降雨熱區的風險趨勢。本研究嘗試解決以下問題：延長高山陣地坡地災害可能肇生的事先預警整備時間、增進對高山陣

地坡地災害事件發生前兆之綜觀環境條件的認識與掌握，以及利用機器學習來量化風險閾值，有效支援強化防災應變指揮官進行高山陣地之坡地災害警戒調度的決心下達。

根據歷史的氣象統計資料，劇烈降雨或累積總降雨量過大常常是導致高山陣地坡地災害的動態因素之一。在坡地災害事件發生前，綜觀環境條件的變化，可能會提供一些前兆信號（林，2022；林，2024）。因此，通過加強對環境條件變化的追蹤監測，可以及時掌握這些參數的變化情況，從而預先發現可能存在的高山陣地坡地災害風險。爰此，落實氣象預報的追蹤監測，持續關注劇烈天氣預報，並運用機器學習瞭解劇烈降雨熱區以及坡地災害的可能變動趨勢，事前預測降雨量及其時空分佈變化。並進一步結合地形地貌特徵，識別易發生滑坡、崩塌或土石流的高風險區域，評估顯著降雨對高山陣地坡地穩定性的影響，強化對累積總降雨量及其時空分佈變化的預測能力，提供精確的降雨預報；通過實時數據更新和動態監測，提前

發出有效預警信號，為我國軍高山陣地在面對氣候變遷劇烈降雨天氣系統影響下，爭取防災與減災提供更多準備時間。

研究內容，將針對高山陣地坡地災害事件進行診斷分析與機器學習，具體執行內容，共計3項，分述說明如下：

- (1) 統計歷史坡地災害事件的氣候特徵、時空分佈與環境條件，旨揭瞭解不同氣候條件、時間和空間尺度下，坡地災害的發生情況和行為特徵，並與事件可能肇因之不同天氣系統進行分類統計，釐清兩者可能存在的因果對應關係(Chang et al., 2013)。
- (2) 利用機器學習建模，建置穩定可靠的高山陣地坡地災害潛勢預報模式。經由交叉驗證挑選後，提出最佳高山陣地坡地災害潛勢前兆之AI方法論，並有效回饋至防災預警的實際整備與操作作業上。
- (3) 建置劇烈降雨天氣系統可能肇致臺灣地區高山陣地坡地災害事件之環境前兆所對應的氣候概念模式，掌握氣象因子以及環境變化趨勢與重要參數閾值，將有助於加強對坡地災害可能肇生的預警和預測能力，整合性的研究分析回饋於防災應變、整備調度與預警警戒等方面，都具有非常重要的參考價值。

2. 研究方法與初步成果

2.1 氣候統計

彙整統計過去18年(2006至2023年)臺灣坡地災害事件的時間、地點、肇因及災情等基礎背景資料，做為本研究的個案樣本(其中，坡地災害事件選取，採以農村水保署官網公告標定之重大災害事件為準)。對應過去18年(2006至2023年)臺灣地區所發生的豪雨個案中，進行劇烈天氣之前兆環境的比對診斷，嘗試釐清可能肇致強降雨之關鍵氣象參數的量化特徵與變化脈動(Weng and Yang,

2018)，並進一步統計分類坡地災害事件可能肇因的天氣系統類型。

所選取坡地災害事件包含有土石流、崩塌、地滑、洪水、沖蝕等。統計分析結果顯示，坡地災害肇生在年際分佈表現上，並無規律變化的特徵(如圖1)，端視該年所影響的劇烈降雨事件而定。至於，前9年的肇生事件數，共計392件，佔總數的67.9%；後9年的肇生事件數，共計185件，佔總數的32.1%，顯見肇生次數近年有減少趨勢。其中，坡地災害事件肇生個數又以2009年最多(共127件)，2020年最少(共3件)，均與颱風肇因有關。最容易肇生的地區分別有新北市、宜蘭縣、南投縣、嘉義縣、高雄市及屏東縣；其中，高屏兩地合佔總數的24.6%最多。另一方面，最常發生的時段落在5至9月，暖季合佔總數的76.9%最多(圖2)，對應颱風與梅雨鋒面等劇烈降雨天氣系統所影響，此一結果與原先預期相當吻合。

然而，我們更關心當有坡地災害事件發生之際，當時的天氣系統以及環境特徵。結果顯示，主要可分為5大類型(圖3)，分述如後：

- (1) 第1種：颱風、熱帶擾動或颱風過後所引進的西南氣流，共計420件，佔總數的72.8%。
- (2) 第2種：梅雨鋒面，共計87件，佔總數的15.1%。
- (3) 第3種：非梅雨鋒面的鋒面，以及午後雷陣雨，共計24件。
- (4) 第4種：東北季風，以及東北季風與南方低壓系統所結合的共伴環流，共計37件，佔總數的6.4%。
- (5) 第5種：地震，共計9件，佔總數的1.5%。

其中，第1種類型加上第2種類型就佔總數的87.9%，由於颱風通常帶來豐沛的水氣和勁風，對地形造成強烈的影響，容易引發土石流、崩塌等高山陣地的坡地災害(Tsai and

Lee, 2008)；梅雨鋒面則常伴隨著長時間的持續性降雨，尤其在持續性暴雨時，也易導致山區土石流、崩塌等坡地災害（林等，2015）。透過統計分析，可以看出颱風（以及熱帶擾動或颱風過後所引進的西南氣流）及梅雨鋒面對我高山陣地坡地災害事件的肇生，扮演著非常關鍵的角色。

2.2 機器學習

土砂災害的種類很多，包括崩塌、土石流、坡面沖蝕等等，肇生的因素眾多且複雜，研究將先選取崩塌類型來進行分析(Lin et al., 2023)，針對2個主要的影響因子，分別為潛在環境因子與動態誘發因子兩類。潛在環境因子又可區分為地形因子、地質因子以及區位因子；另，選取降雨因子做為動態誘發因子來進行綜合診斷。

由於崩塌的發生非常複雜且高度非線性，而由過去的文獻中，類神經網路(Artificial Neural Network, ANN)證實在處理非線性問題上表現優異，Polykretis et al. (2015)證實類神經網路應用於崩塌潛勢分析之可行性。此外，近年來新穎類神經網路多層感知機(Multilayer Perceptron, MLP)和支援向量機(Support Vector Machine, SVM)亦廣泛應用到許多領域，如災害預警等研究，結果均顯示支援向量機SVM在準確度、強健性和訓練模式之速度效率上，皆優於傳統上最常使用的倒傳遞類神經網路。

使用93年至106年衛星判釋全島崩塌地圖，該產品是由林務局委託成功大學，以當年度1至7月全島鑲嵌福衛二號衛星影像（93年度使用SPOT衛星影像），建立自動判釋崩塌地作業標準，繪製崩塌地判釋最小面積為0.1 公頃，各年度統計的崩塌資料筆數如表1，將林務局由衛星影像判釋的崩塌範圍，此資料格式為多邊形向量，轉為網格資料，以利後續模式建立。

擬以臺灣地區近年來坡地災害頻繁的高

屏溪流域作為示範區域，其地理位置如圖4所示。崩塌影響因子資料，選用12個潛在環境因子，分別為坡度、坡向、曲率、平面曲率、剖面曲率、坡長、與河川之距離、與道路之距離、與斷層之距離、高程、地形濕度指數及岩性。動態誘發因子則以降雨水文為主，蒐集研究區域內31個中央氣象局雨量測站資料，由當年度最大累積雨量所引起，分別計算各雨量測站每年最大24小時累積雨量，再以徐昇多邊形法分配各網格的雨量，由於雷達迴波產品(QPESUMS)於山區會有遮蔽效應，導致山區的降雨常常低估，因此使用觀測雨量站資料計算平均降雨。

2.2.1 支援向量機(SVM)

支援向量機(SVM)是一種監督式學習的方法，由Vapnik在1990年代早期發展出來解決統計分類問題，更在1995年將應用範圍擴展到回歸分析的問題上。支援向量機(SVM)主要有兩個的優點：(1)使用結構化風險最小誤差法(structural risk minimization)，有別於傳統類神經使用經驗風險最小化(empirical risk minimization)，支援向量機(SVM)不僅可以降低目標誤差函數，同時也考慮了網路結構複雜度，使模式能達到一定的準確度，又不至於使架構太大而增加運算時間；(2)模式的結構和模式的權重被轉換成一個二次規劃問題(quadratic programming)，可以採用標準演算法快速求解，且能得到全域唯一最佳解，避免傳統方法求解時產生區域最佳解之缺失。

近年來支援向量機(SVM)逐漸在各研究領域拓展，以解決非線性回歸的問題，基本原理為透過訓練資料產生回歸方程式，並在測試資料產生最小誤差之結果，架構圖如圖5所示。支援向量回歸(SVR)的目標為找出一個非線性函數，產生在容許誤差範圍內最近似目標輸出值 y 之網路輸出值 $\hat{y}$ 。考慮 N_d 筆訓練資料 $[(x_1, y_1), (x_2, y_2), \dots, (x_{N_d}, y_{N_d})]$ ，首先，

將輸入向量 $\mathbf{x}$ 經由非線性函數(nonlinear mapping function) 轉換至高維度空間，因此經由輸入向量 $\mathbf{x}$ 產生輸出向量 $\hat{y}$ 的回歸方程式可表示如下：

$$\hat{y} = f(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) + b \quad (1)$$

又基於結構風險最小化原理，推估回歸函數的權重 $\mathbf{w}$ 及偏權值 b ，目標函數如下式：

$$R = \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^{N_d} L_\varepsilon(\hat{y}_i) \quad (2)$$

式中的第一項和第二項分別代表複雜度和經驗誤差，模式複雜度越高其經驗誤差會越小。而此架構的複雜度和經驗誤差間之平衡，取決於懲罰係數 C ，懲罰係數 C 值越大代表經驗誤差對目標函數的影響越大，反之則越小。 L_ε 為誤差函數可表示為：

$$L_\varepsilon(\hat{y}) = |y - f(\mathbf{x})|_\varepsilon = \begin{cases} 0 & \text{for } |y - f(\mathbf{x})| < \varepsilon \\ |y - f(\mathbf{x})| - \varepsilon & \text{for } |y - f(\mathbf{x})| \geq \varepsilon \end{cases} \quad (3)$$

設定懲罰係數為 1，表示使用者視架構的複雜度和經驗誤差兩者同等重要；而容許誤差(tolerance) ε 設定為 0.01。Vapnik(1995)將 SVR 問題轉換成下列的最佳化問題：

Minimize

$$R(\mathbf{w}, b, \xi, \xi') = \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^{N_d} (\xi_i + \xi'_i)$$

subject to

$$y_i - \hat{y}_i = y_i - (\mathbf{w}^T \phi(\mathbf{x}_i) + b) \leq \varepsilon + \xi_i \quad (4)$$

$$\hat{y}_i - y_i = (\mathbf{w}^T \phi(\mathbf{x}_i) + b) - y_i \leq \varepsilon + \xi'_i$$

$$\xi_i \geq 0, \quad i = 1, 2, \dots, N_d$$

$$\xi'_i \geq 0, \quad i = 1, 2, \dots, N_d$$

式中 ξ 和 ξ' 稱為鬆弛變數(slack variables)，兩者分別代表資料點落在容忍誤差範圍外的誤差值，用以說明部分離群的訓練資料。將式(4)之最佳化問題，導入 Lagrange 乘數法(Lagrange multipliers)以 Lagrangian 求解，先轉換成一對偶問題(dual problem)，接著對主要變數($\mathbf{w}, b, \xi, \xi'$)微分後，可整理改寫

如下：

Maximize

$$\sum_{i=1}^{N_d} y_i (\alpha_i - \alpha'_i) - \varepsilon \sum_{i=1}^{N_d} (\alpha_i + \alpha'_i) - \frac{1}{2} \sum_{i=1}^{N_d} \sum_{j=1}^{N_d} (\alpha_i - \alpha'_i)(\alpha_j - \alpha'_j) \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) \quad (5)$$

subject to

$$\sum_{i=1}^{N_d} (\alpha_i - \alpha'_i) = 0$$

$$0 \leq \alpha_i \leq C$$

$$0 \leq \alpha'_i \leq C$$

$$i = 1, 2, \dots, N_d$$

上式為二次規劃問題，求解即可得 α_i 和 α'_i ，每組 α_i 和 α'_i 對應一筆訓練資料，非零的 $(\alpha_i - \alpha'_i)$ 表示該筆資料之估計誤差大於等於 ε ，用於建立目標函數稱之為支援向量(Support Vector)，而為零的 $(\alpha_i - \alpha'_i)$ 則視為多餘可刪除；換言之，支援向量也就是指這些用來協助定義目標函數之資料。一般而言， ε 愈大支援向量就愈少，對解來說，每個支援向量相對重要性就愈高，因此只要使用訓練資料中一個小的子集合即可表示解，在運算上具有相當大的優勢。另外，此最佳化問題的解，由於目標函數為一個凸面(Convex)方程式，因此可得唯一解且為全域最佳解。其中，最佳的 Lagrange 係數 α^* 可利用標準二次規劃演算法求得，回歸函數可以改寫成：

$$f(\mathbf{x}) = \sum_{i=1}^{N_d} \alpha_i^* K(\mathbf{x}, \mathbf{x}_i) + b \quad (6)$$

式中 $K(\mathbf{x}, \mathbf{x}_i)$ 為核函數(Kernel Function)，使用核函數的好處是能處理任何維度的特徵空間，而不必實際進行計算。

$$\text{Linear: } K(x_i, x_j) = x_i^T \cdot x_j \quad (7)$$

Polynomial

$$K(x_i, x_j) = (\gamma \cdot x_i^T \cdot x_j + r)^d, \gamma > 0 \quad (8)$$

Radial basis function:

$$K(x_i, x_j) = (-\gamma ||x_i - x_j||), \gamma > 0 \quad (9)$$

Sigmoid:

$$K(x_i, x_j) = \tanh(\gamma \cdot x_i^T \cdot x_j + r) \quad (10)$$

其中 γ 是 gamma 項、 r 是偏差項 d 為次方項，這三種參數是必須根據不同主題率定的。

最後的回歸公式為

$$f(\mathbf{x}) = \sum_{k=1}^{N_{SV}} \alpha_k K(\mathbf{x}, \mathbf{x}_k) + b \quad (11)$$

式中 $\mathbf{x}_k$ 為第 k 個支援向量； N_{SV} 為支援向量的個數。

隨著現今資訊技術的精進與地理資訊系統的發展，可利用電腦將各種不同空間與時間的圖資進行疊合，方便使用者進行相關規劃作業。由於崩塌潛勢分析為複雜的非線性問題，同時也需考慮大量的地文資料，因此利用地理資訊軟體 ArcGIS 進行歷史崩塌地圖套疊與資料處理，再將處理過的資料透過支援向量機(SVM)進行崩塌潛勢評估模式建置，此模式可計算研究區域內各網格之崩塌潛勢值，即可利用模式分析出坡地災害崩塌潛勢值，繪製高精度與高效能之坡地災害崩塌潛勢地圖。

2.2.1 多層感知機(MLP)

多層感知機(Multilayer Perceptron, MLP)是一種前向傳遞類神經網路，包含三層結構(輸入層、隱藏層和輸出層)，並且利用倒傳遞(Back Propagation)的技術達到學習的監督式學習，由 Rumelhart 等人於 1986 年提出。MLP 由多個的節點層所組成，每一層都全連接到下一層。除了輸入節點外，每個節點都是一個帶有非線性激活函數的神經元(如圖 6 所示)。

MLP 的學習過程分為前饋式及倒傳遞，前饋式學習是指將輸入值由輸入層進入後，經過隱藏層的權重運算，再透過活化函數的轉換，最後在輸出層得到網路輸出值的過程。

$$h_k = f(W_{hx}x_n + b_h) \quad (12)$$

$$y_m = \text{softmax}(W_{yh}y_h + b_y) \quad (13)$$

其中， h_k 為隱藏層第 k 節點單元的輸出值， W_{ij} 為單元 i 連接到單元 j 的權重， x_i 為輸入層第 n 節點單元的輸出值， y_m 為輸出層第 m 節點單元的輸出值， b_i 表示單元 i 的偏差值(Bias)，下標 h 表示隱藏層單元，下標 x 表示輸入層單元，下標 y 表示輸出層單元， $f(\cdot)$ 為激活函數，可以是 sigmoid 函數或 tanh 函數，softmax 函數則為多類別回歸函數。而倒傳遞則是利用最後損失函數(Loss/Cost Function)來進行參數的更新，一般來說都是用誤差均方和(Mean Square Error)當作目標函數。如果誤差值越大，代表參數學習得不好，所以需要繼續學習，直到參數或是誤差值收斂。因此損失函數的優化(Optimization)是機器學習的重要研究目標，最典型的方法是採用隨機梯度下降法(Stochastic Gradient Descent)。當線性關係資料的損失函數為凸函數，可以找到最佳解；然而問題在於非線性關係的資料：其損失函數為非凸函數，求解時容易陷入局部最佳解、而非全域最佳解，則為梯度消失(Vanishing Gradient)，而梯度消失問題會隨著神經網路層數的增加而更加嚴重，意即，隨著梯度逐層不斷消散、導致神經網路對其神經元權重調整的功用越來越小，所以只能轉而處理淺層結構的網路，從而限制了性能。

利用支援向量機(SVM)所建置之崩塌潛勢評估模式，可判斷出各網格之坡地災害崩塌潛勢值，其潛勢值介於 0 到 1，崩塌潛勢值越大代表該地區為崩塌之高危險地區。利用系統性評估後，建議以崩塌潛勢值 0.5 作為門檻值，大於或等於 0.5 視為模式判斷其為崩塌網格；反之，小於 0.5 時模式判斷為穩定狀態(未崩塌網格)。將模式判識之結果與實際觀測值比對，利用混淆矩陣(Confusion

Matrix)判斷各模式之準確率，利用其分類結果，可計算包含、真陽性率(True Positive Rate, TPR)、真陰性率(True Negative Rate, TNR)、偽陽性率(False Positive Rate, FPR)、偽陰性率(False Negative Rate, FNR)以及總準確率(Accuracy, ACC)。其分別定義如下：

TPR：表示歷史記錄之崩塌，模式判別正確的比率。

$$TPR = \frac{TP}{(TP + FN)} \quad (14)$$

TNR：表示歷史記錄之未崩塌，模式判別正確的比率。

$$TNR = \frac{TN}{(TN + FP)} \quad (15)$$

FPR：表示歷史記錄之未崩塌，模式判別成崩塌的錯誤比率。

$$FPR = \frac{FP}{(TN + FP)} \quad (16)$$

FNR：表示歷史記錄之崩塌，模式判別成未崩塌的錯誤比率。

$$FNR = \frac{FN}{(TP + FN)} \quad (17)$$

ACC：表示歷史紀錄之崩塌與未崩塌，模式判別正確的比率。

$$ACC = \frac{TP + TN}{(TP + TN + FP + FN)} \quad (18)$$

TPR 與 TNR 是兩項重要的指標，其值越高越好，在完美判斷的結果時 TPR 與 TNR 兩項指標皆為 1，然而在大部分情形下，TPR 與 TNR 為正負平衡關係。

使用接收者操作特徵曲線(Receiver Operating Characteristic curve, ROC curve)以判斷分類模式優劣。如圖 7 所示，ROC 曲線

是以偽陽性率 FPR 為橫軸，真陽性率 TPR 為縱軸所構成。而在 ROC 曲線下之面積稱之為 AUC(Area Under the ROC Curve)，其優點是其值不受分類樣本數影響，可用來評鑑模式之識別力，其值範圍介於 0 至 1 之間，數值越大，代表分類效果越佳，當 AUC 值超過 0.5，代表模式具有識別能力。

為建置高屏溪流域之坡地災害崩塌潛勢評估模式，提供高山陣地防災預警之參考。首先進行計畫區域資料之蒐集，而後利用 VIF 和 TOL 檢視因子共線性，挑選 VIF<10 且 TOL>0.1 之因子，並使用 IGR 檢視因子的預測能力，挑選 IGR 較高的因子做為模式的輸入項，接著應用支援向量機(SVM)和多層感知器(MLP)方法建模，建模中使用網格搜尋法搜尋各模式的最佳參數，模式表現經過評鑑指標評估及驗證，最後利用所建置之崩塌潛勢評估模式，以此推求研究區域之崩塌潛勢地圖。

進行採樣時，由於崩塌與未崩塌之網格數量差異很大，為避免分析結果偏差，在進行分析時，崩塌資料與未崩塌資料選取之樣本數要相近。因此選取研究區全部的崩塌網格資料，而未崩塌的網格數量大，為避免人為選取主觀印象產生誤差，未崩塌資料選取時，先剔除崩塌網格，再以亂數隨機選取與崩塌資料相同之網格數，單一網格不會被重複選取，一併納入模型中作分析，崩塌與未崩塌網格數比例為 1 比 1。在模式訓練與測試資料部分，由於類神經網路需要較多之訓練資料模式才會精準，因此採用 93 至 99 年資料作為訓練，而 100 年至 106 年為測試資料，訓練與測試資料之比例約為 2 比 1，由以上訓練和測試資料和分析之輸入因子建模，再以混淆矩陣等評鑑指標評估模式表現，最後繪製崩塌潛勢圖做為防災單位決策依據。

採用 12 個潛在環境因子分別搭配 24 小時累積雨量，以支援向量機(SVM)和多層感

知器 MLP 建置兩個崩塌潛勢評估模式。在因子篩選的過程中，使用 IGR 檢定因子的預測能力，接著使用 TOL 和 VIF 檢定分析崩塌因子間的共線性。由表 2 混淆矩陣結果可知 MLP 在訓練階段，其模式判斷崩塌之準確率 (TPR) 可高達 85.7%，以及總準確率 (ACC) 為 79.4%，表示模式在崩塌判斷上具有良好的判斷能力。MLP 在測試階段也能維持良好的表現，實際崩塌資料共 24,897 筆，模式成功模擬出 18,602 筆，結果顯示準確率高達 74.7%，表示模式在辨別崩塌上具有高度的判釋能力，模式總準確率也高達 73.5%。此外，MLP 在測試階段 ROC 曲線下面積 AUC 值為 0.815，顯示模式具有良好之識別能力 ($AUC > 0.5$)。再將模式應用到完整研究區域，將分析結果與實際觀測值相比較，如圖 8 展示 99 年訓練結果應用、圖 9 和圖 10 展示 103 和 106 年訓練結果應用所示，以利瞭解模式應用在完整研究區域表現之優劣。可以發現 2011 年新增之崩塌地多分布在東半部山區，而研究所模擬之結果圖 8 (b) 及圖 8 (c) 顯示模式在研究區域西半部平原地區，能有效反映出研究區域內未崩塌地區，而模式分析出坡地災害崩塌潛勢值高的地區主要分布在東半部山區，與實際觀測值之崩塌分布趨勢相同，圖 9 和圖 10 也具有相同趨勢，此外 MLP 和 SVM 兩模式判勢崩塌的範圍與實際觀測值相似，但模式對崩塌潛勢值的推估略有不同，如 MLP 對推估高潛勢區的範圍較大 (0.9~1)，而 SVM 對於中高潛勢 (0.7~0.9) 的推估範圍較廣，此現象也反映在 TPR 的值中，MLP 的測試資料 TPR 為 74.7%，而 SVM 的 TPR 僅 69.9%。

依據混淆矩陣結果顯示，MLP 和 SVM 在判釋崩塌上之準確率皆約為 70%，而總準確率也能達到 72% 以上。此外，MLP 和 SVM 之 ROC 曲線下面積 AUC 值在測試分別為 0.815 以及 0.794，表示模式具有良好識別能

力 ($AUC > 0.5$)。整體而言，潛勢圖展示中 MLP 或 SVM 沒有明顯的差異，而由 TPR 和 AUC 均顯示 MLP 表現比 SVM 佳，建議未來可使用 24 小時累積雨量 (一日暴雨) 配合 MLP 模式進行崩塌潛勢評估。由上述分析結果可知，無論由評鑑指標或崩塌潛勢圖模擬，均顯示研究所建置之模式具有良好之模擬能力，是以可利用此高山陣地崩塌潛勢評估模式，分析不同雨量情況下之崩塌潛勢圖，以判斷崩塌可能發生之區域，以利官兵進行防災準備。

3. 結論與建議

透過氣象統計分析結果顯示，在整體的 18 年研究母體樣本中，前 9 年的肇生事件數，共計 392 件，佔總數的 67.9%；後 9 年的肇生事件數，共計 185 件，佔總數的 32.1%，顯見肇生次數近年有減少趨勢。其中，坡地災害事件肇生個數又以 2009 年最多 (共 127 件)，2020 年最少 (共 3 件)，均與颱風肇因有關。

研究評估，第 1 種類型加上第 2 種類型就佔總數的 87.9%，由於颱風通常帶來豐沛的水氣和勁風，對地形造成強烈的影響，容易引發土石流、崩塌等高山陣地的坡地災害；而梅雨鋒面則常伴隨著長時間的持續性降雨，尤其在持續性的暴雨時，也易導致山區土石流、崩塌等坡地災害。透過統計分析，可以看出颱風 (以及熱帶擾動或颱風過後所引進的西南氣流) 及梅雨鋒面對我高山陣地坡地災害事件的肇生，扮演著非常關鍵的角色。

再從高山陣地崩塌潛勢評估，需要考慮地形、地質、雨量等不同因素，並包含時間與空間之變異性及非線性因子，研究利用處理非線性問題中表現良好之方法支援向量機 (SVM) 及多層感知器 (MLP) 建立高屏流域崩塌潛勢評估模式。選用 12 個潛在環境因子外，為進一步瞭解降雨對崩塌之影響，加入

24 小時累積雨量作為誘發因子，俾利進行模式建置與討論。

探討降雨所造成之淺層崩塌，蒐集 12 個潛在環境因子和 24 小時最大累積降雨，並以方差膨脹因子(VIF)和容忍度(TOL)量化崩塌因子的共線性，且以資訊增益比(IGR)量化因子的預測能力，篩選崩塌影響因子，模式經過測試與評估後，映射不同重現期距雨量之高山陣地崩塌潛勢圖。

使用 93 年至 106 年衛星影像判釋之崩塌資料，以 101 年至 106 年資料進行模式驗證，透過混淆矩陣和 ROC 曲線下面積(AUC)評估模式之表現以確立此方法之可行性。結果顯示，共線性分析和剔除掉共線性高且預測能力低的高程、坡長、曲率和平面曲率。由多層感知器 MLP 的建置之崩塌潛勢評估模式模擬崩塌之準確率(TPR)高達 74.7%，總準確率(ACC)為 73.5%，AUC 值高達 0.815，顯示建置此模式對於高山陣地坡地災害崩塌潛勢分析有優異之成果。使用採樣並配合評鑑指標以評估模式表現，透過所建置之模式分析完整研究區域之高山陣地崩塌潛勢圖，藉以探討模式在各年度大範圍區域之掌握能力，結果亦顯示模式能精準的模擬出崩塌分布的趨勢。

短期目標將朝綜觀環境氣象量場的運動過程，進行更高解析度的時空尺度診斷分析，瞭解變化的熱力及動力物理機制以及強度演變，建置劇烈降雨個案環境的基礎資料庫，精進對高山陣地坡地災害前兆的關鍵訊號掌握，並在實際守視監測作業上，進行數據量化的校驗修正。中、展期目標，藉以本研究建置之高山陣地崩塌潛勢評估模式，分析研究區域中各網格之崩塌潛勢值，建立崩塌潛勢地圖，讓應用者透過此圖層提前判斷易發生崩塌之地區，俾利進行後續防災之規劃。未來，將納入深度學習模式，精進崩塌潛勢

評估模式模擬崩塌之準確率，並可再行搭配即時的24小時累積降雨資訊（如QPESUMS），有效產製崩塌潛勢圖，提供防災人員決策運用的參考資訊。

4. 參考文獻

- Chang, C.-P., Y.-T. Yang, and H.-C. Kuo, 2013: Large increasing trend of tropical cyclone rainfall in Taiwan and the roles of terrain. *J. Climate*, 26, 4138–4147.
- Lin Te-En , Ming-Jui Chang, Jihn-Sung Lai ,2023 : Environmental diagnosis of precursor of slope hazard potential, 2023 ICEO&SI Conference, June 12-13, Taipei, Taiwan.
- Tsai, H.-C. and Lee,T.-H.,2008: Maximum covariance analysis of typhoon surface wind and rainfall relationships in Taiwan. *J.Appl. Meteor. Climatol.*, 48,997-1016.
- Weng, S.-P., and C.-D. Yang, 2018: The construction and verification of daily gridded rainfall dataset (1960–2015) in Taiwan (in Chinese with English abstract). *Taiwan Water Conserv.*, 66, 33–52.
- 李鎮洋、賴文基、陳振宇、黃效禹、郭力行，2011：莫拉克颱風複合型災害發生歷程的時空重建——以小林村深層崩塌為例，*中華水土保持學報*，42-4，313-324。
- 林冠瑋、顏心儀、郭賢立，2015：豪雨觸發之大規模山崩的降雨條件分析，*工程環境會刊*，35，1-16。
- 林得恩，2022：劇烈降雨系統對高山陣地之風險評估，*氣象預報與分析*，250，1-15
- 林得恩，2023：梅雨鋒面肇致坡地災害之先兆環境研究，*氣象預報與分析*，252，1-15。

5. 表附錄

表 1 各年度崩塌資料筆數

民國	衛星判釋高屏溪流域 崩塌地網格數
93	55263
94	150665
95	52862
96	41331
97	53221
98	21904
99	492854
100	120935
101	75425
102	46891
103	72954
104	46568
105	23748
106	76868

表 2 MLP 模式訓練(測試)結果之混淆矩陣

實際值 \ 模擬值			準確率 (%)
	崩塌	未崩塌	
崩塌	132195 (18602)	41541 (6900)	85.7 (74.7)
未崩塌	22140 (6295)	112794 (17997)	73.1 (72.3)
ACC = 79.4% (73.5%); AUC = 0.873 (0.815)			

註：總準確率(ACC)、ROC 曲線下的面積(AUC)

6. 圖附錄

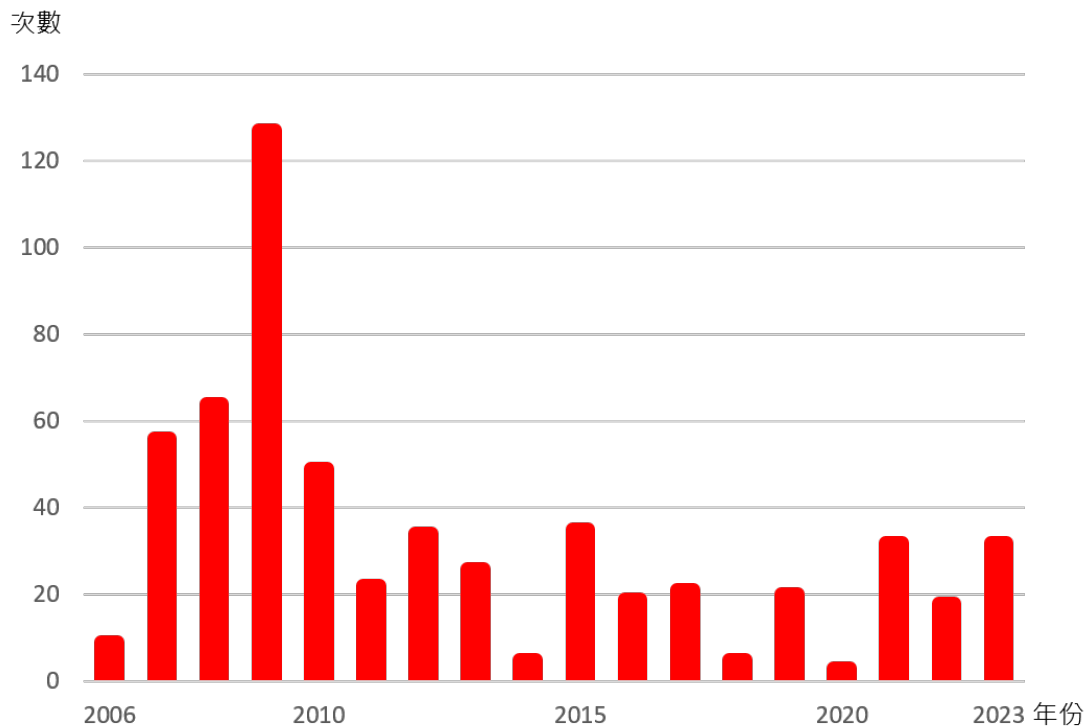


圖 1 2006 至 2023 年期間，逐年坡地災害事件肇生次數統計。

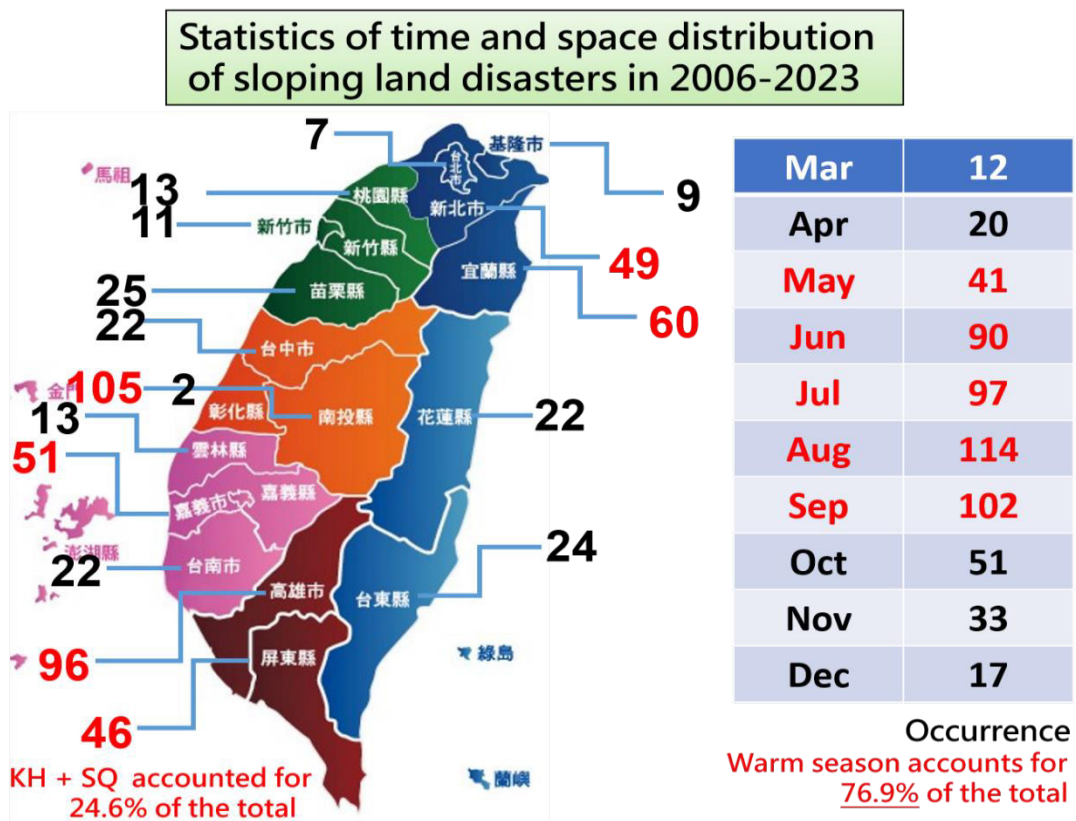


圖 2 2006 至 2023 年期間，坡地災害事件肇生的時空分佈統計。

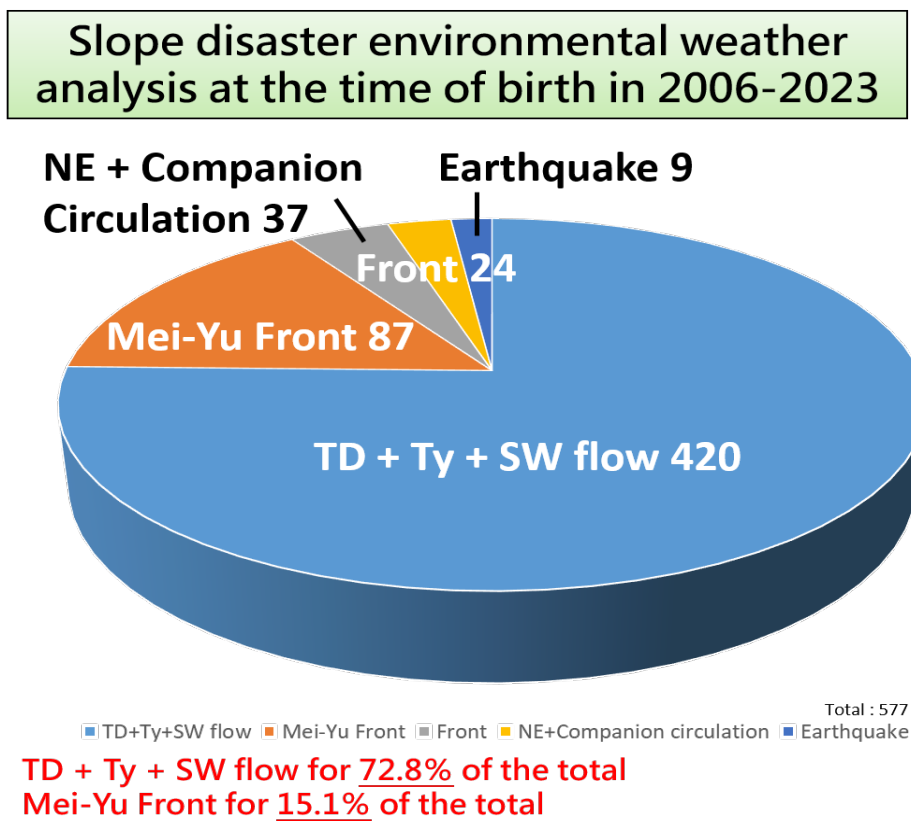


圖 3 2006 至 2023 年期間，坡地災害事件肇因的統計分析。

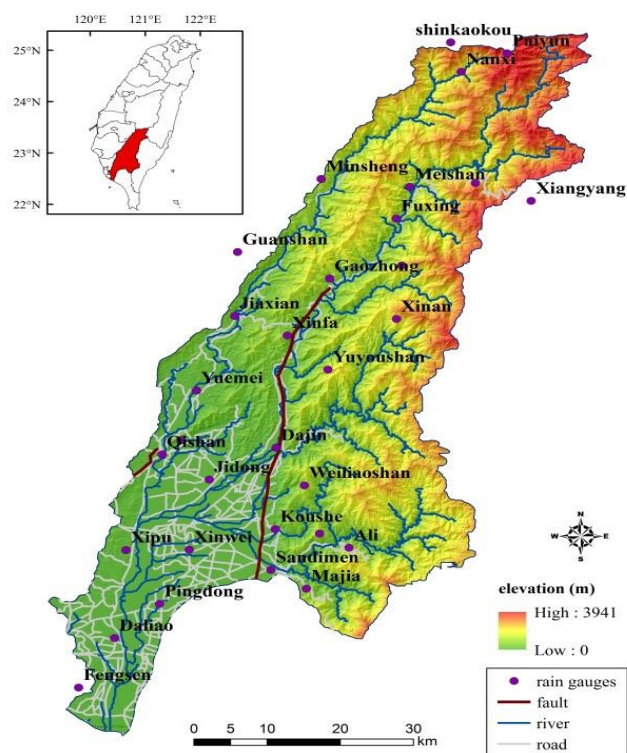


圖 4 計畫研究區域

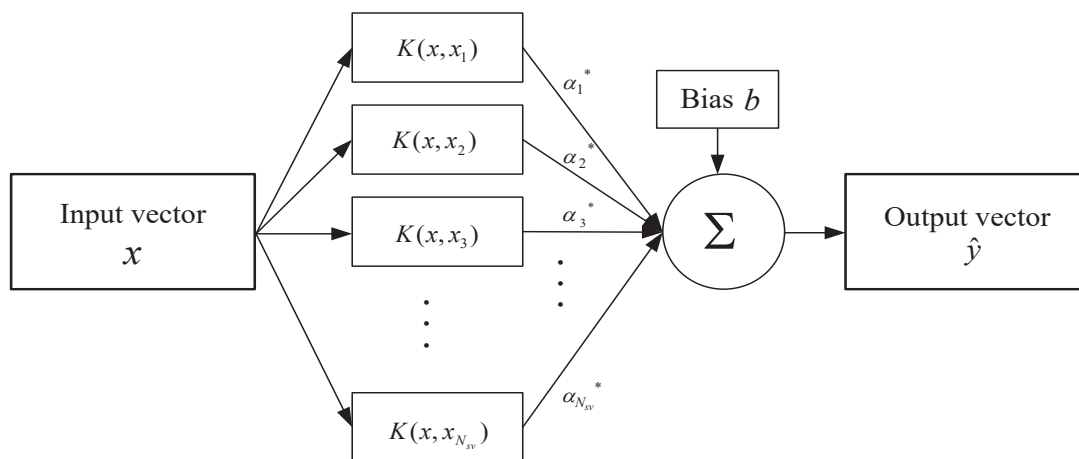


圖 5 支援向量機架構圖

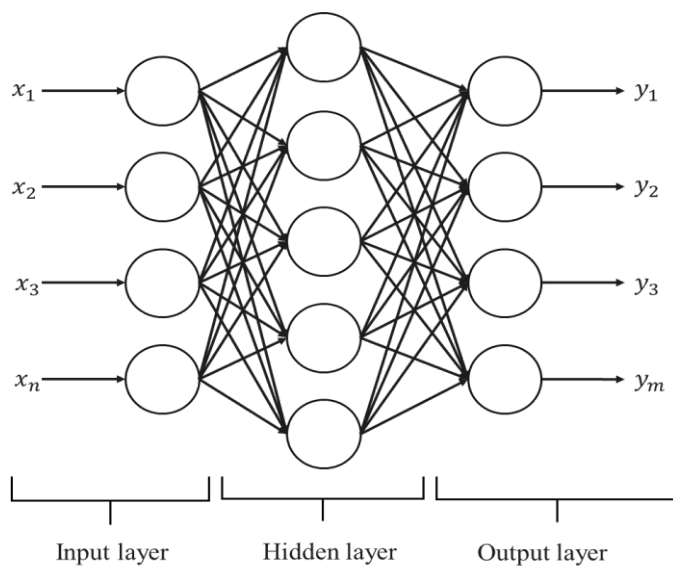


圖 6 MLP 架構圖

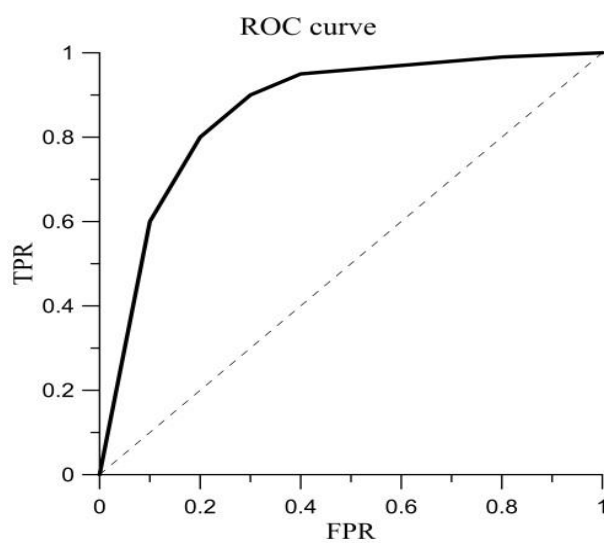
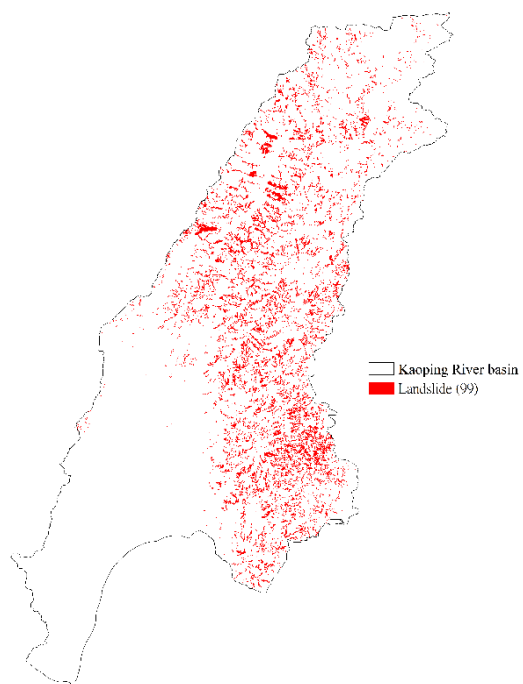
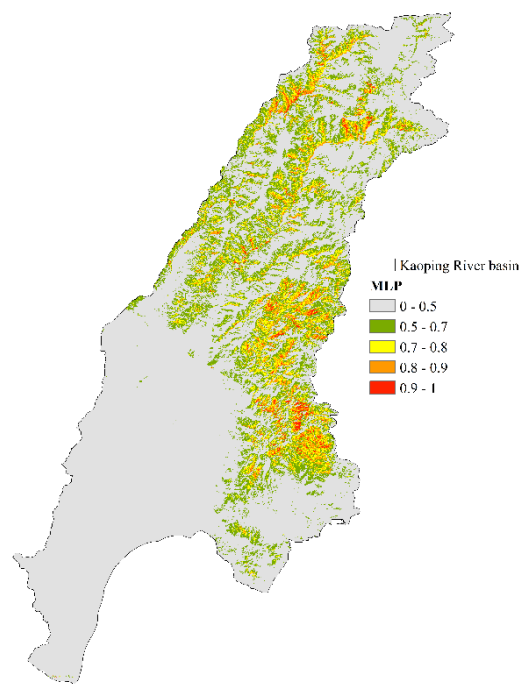


圖 7 接收者操作特徵曲線示意圖

(a)



(b)



(c)

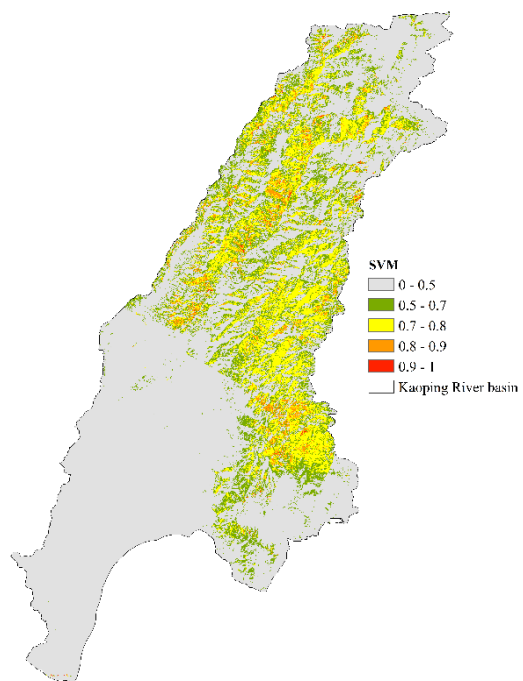


圖 8 99 年訓練資料展圖。(a)新增崩塌地、(b) MLP 模式產製的崩塌潛勢圖、(c) SVM 模式產製的崩塌潛勢圖。

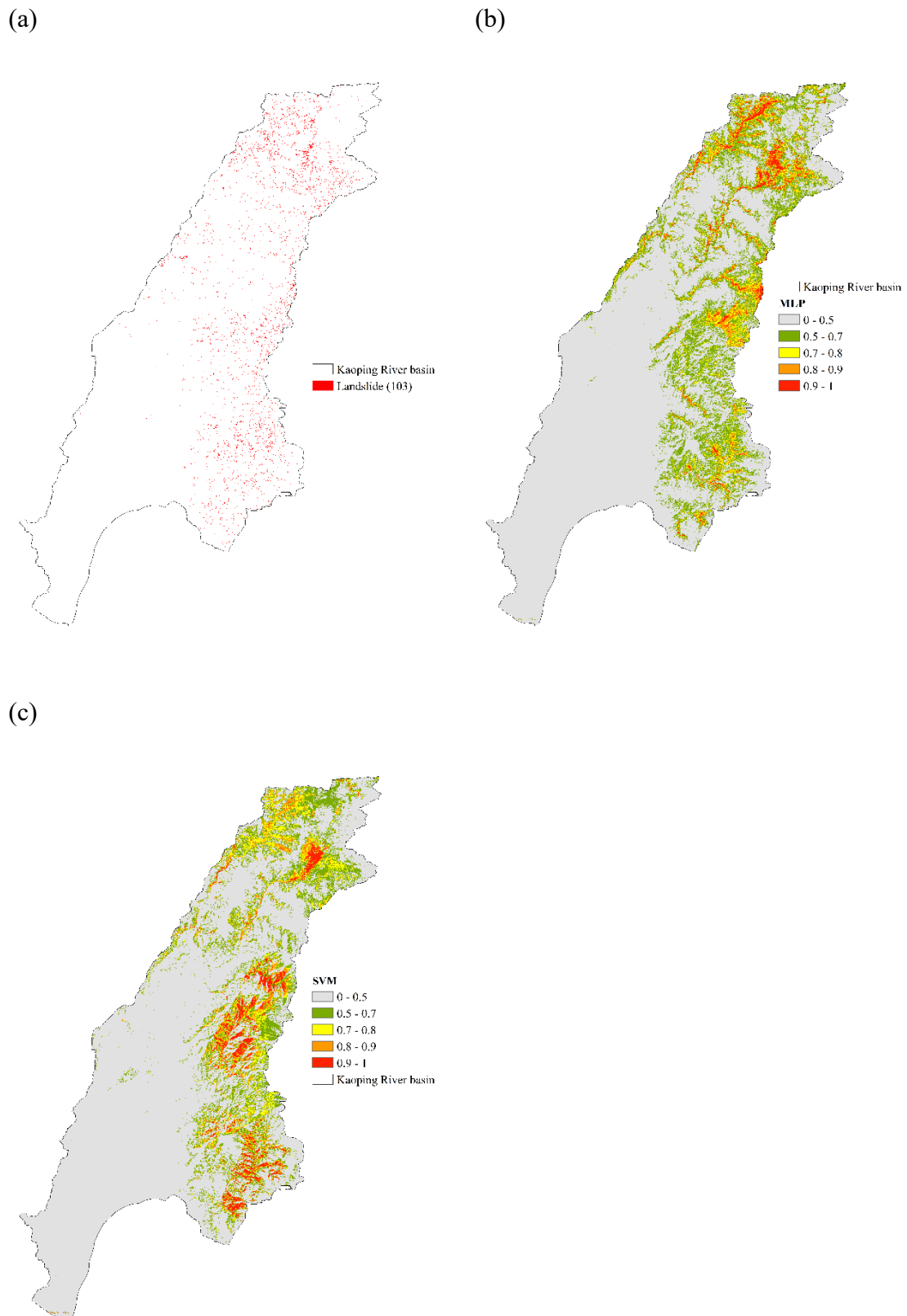
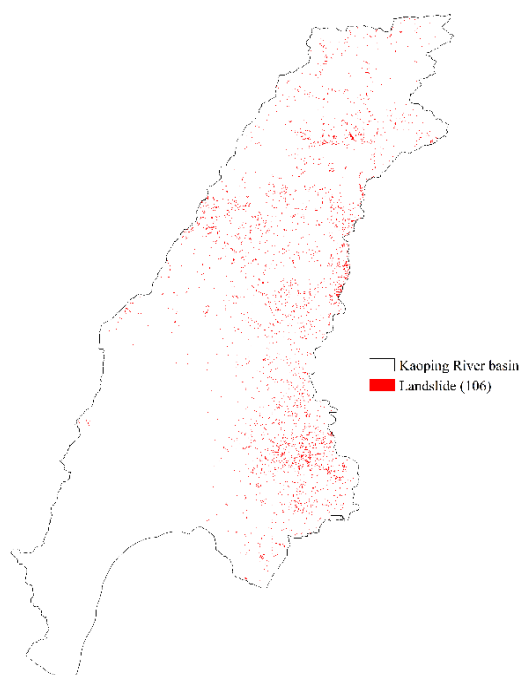
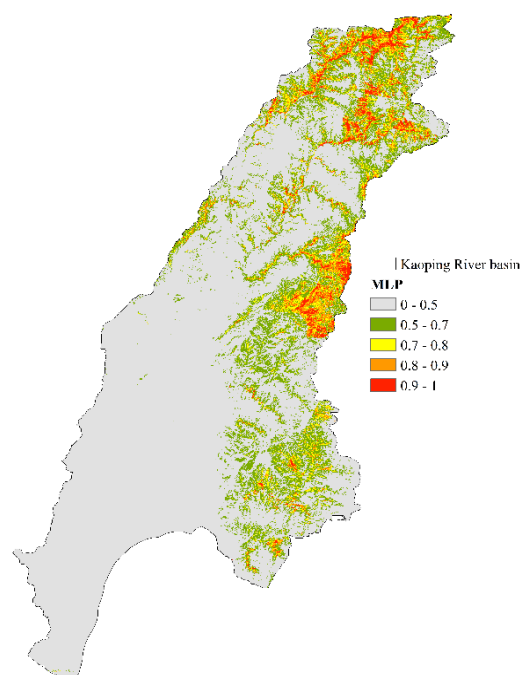


圖 9 103 年測試資料展圖。(a)新增崩塌地、(b) MLP 模式產製的崩塌潛勢圖、(c) SVM 模式產製的崩塌潛勢圖。

(a)



(b)



(c)

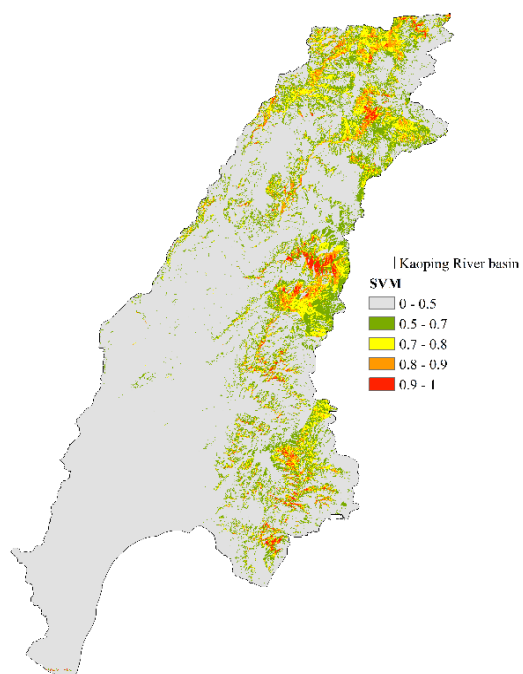


圖 10 106 年測試資料展圖。(a)新增崩塌地、(b)MLP 模式產製的崩塌潛勢圖、(c)SVM 模式產製的崩塌潛勢圖。

Research on the Application of Machine Learning in Diagnosis and Early Warning of Severe Rainfall Events

De-En Lin, Ming-Jui Chang

Center for weather and climate disaster research, NTU

Abstract

Our military's high mountain positions are faced with uneven spatial and temporal distribution of rainfall caused by weather systems, as well as record-breaking total cumulative rainfall. Slope disasters will be even more severe tests and challenges in the future.

The main purpose of this study is to diagnose and analyze historical slope disaster events through scientific and quantitative research and judgment steps, thereby clarifying the characteristics of key meteorological parameters of the environment at that time, thereby extending the advance warning and preparation time for possible slope disasters, and effectively The decision was made to support and strengthen disaster prevention and response commanders in disaster alert and dispatch.

Using research methods such as climate statistics and machine learning, a case study was conducted on severe rainfall in Taiwan. Use machine learning modeling to build a stable and reliable collapse potential prediction model, propose an AI methodology for the best collapse potential precursors, and effectively feedback to the actual preparation operations of the army's disaster prevention and early warning.

.

Keywords: Alpine Positions, Slope Disasters, Severe Rainfall Events