

# 資料探勘決策樹方法應用於閃電躍升預警臺灣午後對流強降雨之研究

葉明生<sup>1</sup> 陳新淦<sup>2</sup> 林秉煜<sup>2</sup> 馮欽賜<sup>2</sup> 黃椿喜<sup>2</sup> 劉承翰<sup>1</sup> 鄭安孺<sup>1</sup>

多采科技有限公司<sup>1</sup>  
交通部中央氣象局<sup>2</sup>

## 摘要

過去已有文獻對於臺灣地區的閃電及降雨進行研究，指出劇烈天氣常伴隨閃電發生，而閃電頻率高的地區經常伴隨強烈對流及降雨。本研究初步採用2017年各午後對流事件，以每0.05°解析度為計算的網格中心，且取此中心周圍半徑5公里範圍作為閃電的配置及觀測降雨的計算(使用QPESUMS雷達降雨估計資料)，採用Gatlin(2006)閃電躍升演算法計算閃電躍升。而決策樹(Decision Tree)是資料探勘(Data Mining)中常使用的方法之一，具特徵萃取與描述功能，可用來分析與預測。為強化閃電躍升對於短延時強降雨之預警，本研究以60分鐘累積雨量大於40 mm作為大雨預警的基準，由閃電數量、閃電躍升數及已發生降雨量所組合的多種已發生條件因子納入決策樹分析，探討閃電躍升搭配其他因子提升臺灣午後對流強降雨之預警成效。

經初步分析，在僅以10分鐘降雨量分析後續發生大雨機率的條件情況，當下10分鐘雨量若達到10 mm以上，至未來50分鐘有31%機率會發生大雨；而只以發生閃電躍升為條件，則約有50%機率發生大雨。運用決策樹整合上述二條件之結果顯示，以連續閃電躍升數，搭配10分鐘雨量(或20分鐘雨量)，可提高大雨發生機率判斷，當發生閃電躍升且10分鐘雨量大於10 mm，於未來60分鐘達大雨機率大約為70%。因此，本研究結果呈現閃電躍升配合已累積10分鐘雨量的資訊，可作為強化臺灣午後對流強降雨預警作業之參考。

中文關鍵字：決策樹;閃電躍升;午後對流

## 一、前言

強對流雷雨胞的對流發展過程中經常伴隨閃電與降水的發生，雷雨胞活躍期間的閃電活動會快速增加，於總閃電率達峰值前，閃電次數通常會呈現急遽增加的現象，此現象稱為閃電躍升(Lightning jump)(Williams et al. 1999)。閃電躍升現象的客觀判定方法以及其與局部地區發生劇烈天氣(包括龍捲風、地面強風及冰雹)事件的關係，在國外研究有Williams et al. (1999)針對發生於1996至1998年美國佛羅里達地區的劇烈雷暴(伴隨劇烈天氣之雷暴)事件個案，發現劇烈雷雨事件的總閃電率皆超越60 fpm (flashes/min；每分鐘的閃電次數)，而非劇烈雷雨的總閃電率多為介於1到10 fpm，此外，總閃電率峰值出現的時間相較於地面觀測到的劇烈天氣事件約有5~20分鐘的領先。Goodman et al. (2005)利用位於阿拉巴馬州超高頻LMA (Lightning Mapping Array)閃電偵測系統所得的資料，研究2002年夏季強對流以及龍捲風暴風雨事件中的總閃電行為。近期則有Gatlin (2006)、Schultz et al. (2009)、Gatlin and Goodman (2010)陸續提出客觀和定

量的閃電躍升演算法，分別應用於不同劇烈天氣預警。

國內相關研究有洪(2002)利用雲對地閃電(CG)分析臺灣北部山區之午後對流個案與雷達回波的關係。譚(2006)分析1989~2005年臺灣颱風和對流系統之閃電與降雨的特徵，發現閃電頻率於中央山脈西部最為密集。林等(2006)以2005年5月12日梅雨鋒面個案，分析大部份的閃電發生在較強降水區，但傾向於偏向降雨中心的前緣，閃電與降雨兩者在空間及時間上的分佈相似。戴等(2015)利用台電公司全閃電偵測系統的IC資料，發現IC的躍升對午後雷暴、移動性雷暴及颱風雨帶環境下的移動性雷暴三種不同型態雷暴系統，降雨領先時間分別約有15~40分鐘。戴等(2017)使用台電公司全閃電偵測系統的雲中閃電資料，顯示IC躍升對於臺灣對流性降雨有預警應用價值。葉等(2017)使用台電公司與中央氣象局建置的閃電偵測系統所得觀測資料，探討適用臺灣地區閃電躍升演算法與參數。

決策樹(Decision Tree)是資料探勘(Data Mining)中常使用的方法之一，具特徵萃取與描述功能，可用來分析與預測。為強化閃電躍升對於短延時強降雨之預警，本研究以60分鐘累積雨量大於40 mm作為大雨

預警的基準，由閃電數量、閃電躍升數及已發生降雨量所組合的多種已發生條件因子納入決策樹分析，探討閃電躍升搭配其他因子提升臺灣午後對流強降雨之預警成效。

## 二、閃電與降雨資料來源與解析

### (一)中央氣象局閃電偵測系統(Central Weather Bureau Lightning Detection System, CWBLDS)

本研究採用氣象局閃電資料，氣象局於2015年4月開始於臺灣地區建置閃電偵測系統，氣象局的閃電偵測資料自2015年4月開始提供，並已整合至氣象局官網展示系統以及WINS系統中，至2019年底氣象局已有的26個閃電偵測系統測站，分佈如圖1所示。使用儀器為TOA公司(TOA Systems, Inc.)的閃電偵測設備，儀器感應敏感的波段主要是位於低頻的電磁波，因此其閃電偵測類型主要是雲對地閃電觀測。CWBLDS主要使用到達時間差法(Time Difference of Arrival, TDOA)定位閃電來源。

### (二)網格化雨量資料

本研究蒐集QPESUMS劇烈天氣偵測系統逐10分鐘網格時雨量估計產品，目前QPESUMS系統是每10分鐘產出一組網格化時雨量估計資料，因此需要另行計算10分鐘累積雨量。因有每10分鐘一筆的小時累積降雨量，則採用一小時累積降雨量來推算10分鐘降雨量。而缺QPESUMS降雨資料的時間，將以克利金法內插地面測站雨量觀測補遺。

### (三)資料集整

逐閃電網格分析各網格2017年全年閃電躍升時間序列。分析方法為取網格中心周圍半徑5公里範圍內的氣象局CWBLDS總閃電資料，以Gatlin (2006)閃電躍升演算法，計算閃電躍升，時間解析度為1分鐘。

由於本研究之分析目標為午後對流系統，因此先依豪大雨天氣特報挑選午後對流事件，再經人工比對排除其中因其他天氣系統引發降雨事件，得出2017年午後對流事件。

## 三、午後對流10分鐘降雨量後續發生大雨之機率分析

本研究先分析2017年60天午後對流日10分鐘降雨量對應發生大雨之機率。表1是各雨量站在60天午後對流日中的11~19時，第1次符合「10分鐘降雨量 $\geq 6\text{mm}$ 」、...、「10分鐘降雨量 $\geq 40\text{mm}$ 」之發生大雨之機率。其計算方式與各欄位說明，以「10分鐘降雨量 $\geq 6\text{mm}$ 」為例，說明如下：

當某站於各午後對流日的11~19時，第1次符合「10分鐘降雨量 $\geq 6\text{mm}$ 」的時間稱為 $t_0$ ，而第1次符合「10分鐘降雨量 $\geq 6\text{mm}$ 」之未來10分鐘稱為 $t_{10}$ ，以此類推，第1次符合「10分鐘降雨量 $\geq 6\text{mm}$ 」之未來50分鐘稱為 $t_{50}$ ，如圖2所示。

接下來，則先計算當日11時至 $t_0$ ，計算每逐10分鐘之時雨量是否有達到大雨，若是，則計算在「已發生大雨(含 $t_0$ )」的發生大雨之機率；計算當日 $t_0$ 至 $t_{10}$ 是否有達到大雨，若是，則計算在「未來10分鐘」有發生大雨。並依序分別計算至 $t_{20}$ 、...、至 $t_{50}$ ，是否達大雨標準。如圖2，於 $t_0$ 未達大雨標準，至 $t_{10}$ 也未達大雨標準，但 $t_{20}$ 時雨量 $42\text{mm}$ 達大雨，故至未來20、30、40、50分鐘，皆為達大雨標準。如此，即可統計「已發生大雨(含 $t_0$ )」、至未來10分鐘、...、至未來50分鐘的發生大雨之機率。最後一欄，則為當日至未來50分鐘，皆未達大雨。

如表1所示，在10分鐘雨量達到 $10\text{mm}$ 以上，至未來50分鐘只有30.9%機率會發生大雨，但在10分鐘雨量達到 $14\text{mm}$ 以上，至未來50分鐘有50.4%機率會發生大雨。

## 四、在已發生之條件以決策樹分析達降雨門檻機率

決策樹(Decision Tree)是機器學習(Machine Learning)中預測模型之一，也是資料探勘(Data Mining)中常使用的方法之一，具有監督式之特徵萃取與描述功能，同時可用來分析與預測，與其他模型相較之下，運算速度及可解釋性(Interpretability)為其優勢，本研究採用連續閃電躍升數為前提，以決策樹配合其它已知條件，建立達降雨門檻值之發生機率。

採用閃電與降雨資料為2017年各午後對流事件，以閃電計算網格為中心，閃電與QPESUMS降雨搜尋範圍皆採用建議之半徑 $5\text{km}$ ，計算閃電躍升並取出第1個達到降雨門檻前之閃電躍升進行分析。10分鐘降雨量以各閃電網格中心位置，於半徑 $5\text{km}$ 範圍內取QPESUMS網格中雨量最大者為代表。

以圖3為例，說明如何取出連續閃電躍升數，以供後續決策樹分析。在圖3中，於第1次達降雨門檻值( $40\text{mm/hr}$ )之前共3個閃電躍升群組(在考量預警的時間需求，若二個閃電躍升之間隔小於等於1分鐘，則視為同一個閃電躍升群組)，考量即時作業，連續時間閃電躍升會持續多久是未知的。因此，站在每個閃電躍升當下，可分別取出連續躍升數=1有3筆；連續躍升數=2有2筆；連續躍升數=3有1筆。

經上述方式，以達大雨 $40\text{mm/hr}$ 為降雨門檻值，共取出27,941筆連續閃電躍升，其中於未來60分鐘後達大雨有15,249個(比例佔0.546)，未達大雨有12,692個(比例佔0.454)，當發生閃電躍升，約有一半機率發生大雨。決策樹分析，採用scikit-learn 套件，以python程式撰寫，考量因子有閃電躍升、閃電數及已發生降雨量三大類，共20個。

在分別以連續閃電躍升數為1,2,...,及 $\geq 6$ ，隨機挑選70%筆數進行訓練之決策樹，由各決策樹主要的分類條件是10分鐘雨量(門檻約 $4\text{--}5\text{mm}$ 、 $10\text{mm}$ )與20分鐘雨量(門檻約 $17\text{mm}$ )，而與閃電數相關因子，只有

「連續躍升期間之平均分鐘閃電數」出現於連續閃電躍升數 $\geq 6$ 之第2層的分類條件。而依據決策樹之分類條件，可依達大雨機率由低至高，區分為分類1、分類2及分類3等3類。

圖4為以連續閃電躍升數等於1為前提之決策樹(降雨門檻值為40mm/hr)，其分類筆數與達降雨門檻值比例，分別如表2，於連續閃電躍升數等於1，分類1達大雨比例僅為0.218，分類2達大雨比例為0.412，分類3達大雨比例提高為0.751，且隨著連續閃電躍升數增加，分類3達大雨機率明顯增加。

以降雨門檻值為40mm/hr，採用連續閃電躍升數為前提，將決策樹之各分類達大雨機率與平均領先時間，繪製於圖5，此圖除了可提供各分類達大雨機率，亦標示對應平均領先時間(計算方式為各分類達大雨筆數之領先時間取平均值)。

由圖5，就分類1而言，達大雨機率由0.217增加至0.584，但平均領先時間由34.2分鐘減少為20.1分鐘；分類3而言，達大雨機率由0.751增加至0.916，但平均領先時間由14.1分鐘減少為12分鐘，可看出隨著連續閃電躍升數增加，各分類達大雨機率增加，但相對的平均領先時間卻減少。

## 五、結論與建議

午後對流10分鐘降雨量後續發生大雨之機率分析，10分鐘雨量達到10mm以上，至未來50分鐘有30.9%機率會發生大雨。而只以發生閃電躍升為條件，約有一半機率發生大雨。整體而言，以連續閃電躍升數，搭配10分鐘雨量(或20分鐘雨量)，可提高大雨發生機率判斷：當發生閃電躍升且10分鐘雨量 $>10\text{mm}$ ，達大雨機率大約0.7；當發生閃電躍升且10分鐘雨量 $\leq 4\sim 5\text{mm}$ ，達大雨機率大約只有0.2。建議後續可納入更多年份資料，與氣象因子，配合決策樹分析。

## 參考文獻

- 林品芳、張保亮、丘台光、陳嘉榮，2006：閃電資料在降水估計之初步分析，天氣分析與預報研討會。
- 洪景山，2002：雲對地閃電和雷達回波參數之相關：個案研究。大氣科學，30，1，21-34。
- 戴志輝、王尹懋、王安翔、林博雄，2015：雲中閃電資料應用於雷暴即時預警之研究。大氣科學，43，2，115-132。
- 戴志輝、王尹懋、楊明仁、林博雄，2017：雲中閃電預警對流性降雨初探。大氣科學，45，1，43-55。
- 葉明生、張博雄、陳新淦、李天浩、于芃，2017：適用於台灣地區降雨預警之閃電躍升演算法探討，大氣科學，45，4，349-374。
- 譚振威，2006：臺灣地區閃電與降雨的分類及其氣候特徵。國立中央大學大氣物理研究所碩士論文，120頁。

- Gatlin, P. N., 2006: Severe weather precursors in the lightning activity of Tennessee Valley thunderstorms. M.S. thesis, The University of Alabama in Huntsville, 87 pp.
- Gatlin, P. N., and S. J. Goodman, 2010: A total lightning trending algorithm to identify severe thunderstorms. J. Atmos. Oceanic Technol., 27, 3–22.
- Goodman, S. J., R. Blakeslee, H. Christian, W. Koshak, J. Bailey, J. Hall, E. McCaul, D. Buechler, C. Darden, J. Burks, T. Bradshaw, and P. Gatlin, 2005: The North Alabama Lightning Mapping Array: Recent severe storm observations and future prospects. Atmos. Res., 76, 423–437.
- Schultz, C. J., W. A. Petersen, and L. D. Carey, 2009: Preliminary development and evaluation of lightning jump algorithms for the real-time detection of severe weather. J. Appl. Meteor. Climatol., 48, 2543–2563.
- Williams, E.R., B. Boldi, A. Matlin, M. Weber, S. Hodanish, D. Sharp, S. Goodman, R. Raghavan, and D. Buechler, 1999: The behavior of total lightning activity in severe Florida thunderstorms. Atmos. Res., 51, 245–265.

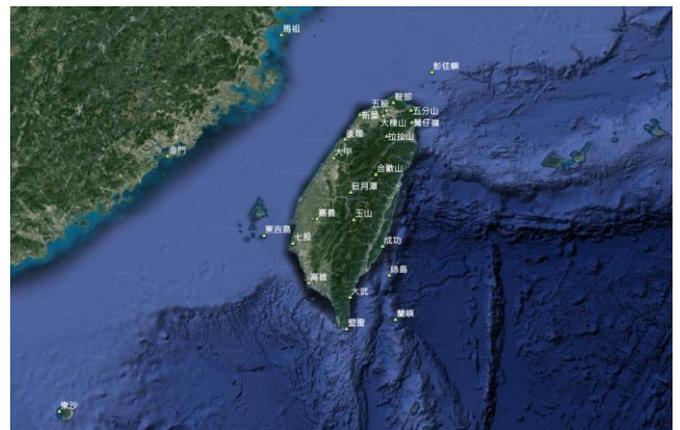


圖1 中央氣象局閃電偵測系統測站分布位置

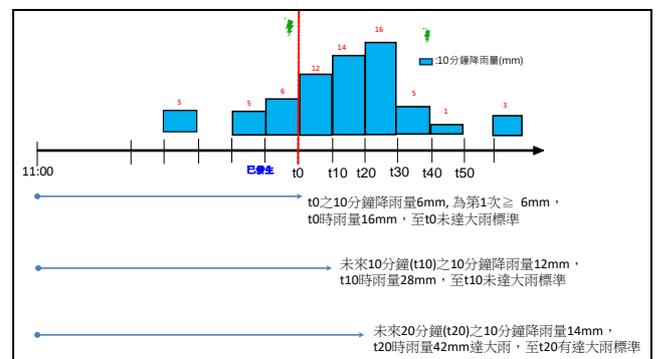


圖2 午後對流10分鐘降雨量對應發生大雨計算之示意圖(以第1次符合「10分鐘降雨量 $\geq 6\text{mm}$ 」為例)

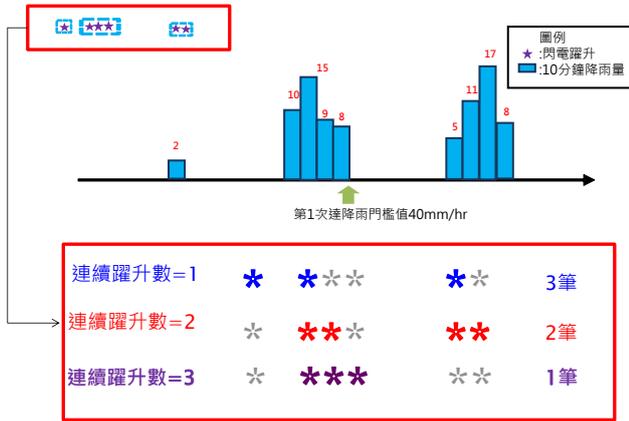


圖3 採用連續閃電躍升數為前提之決策分析資料示意圖

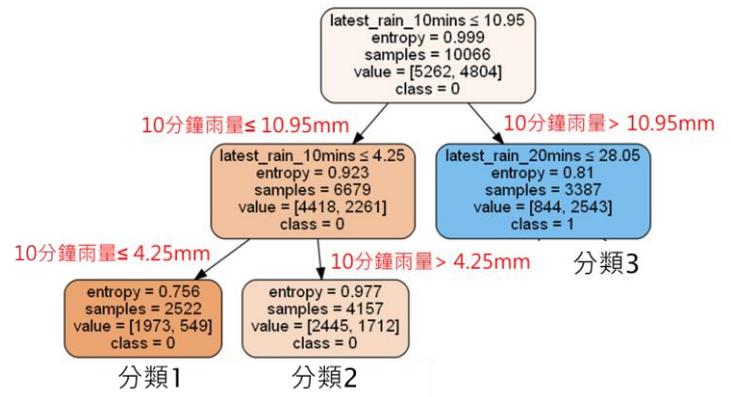


圖4 以連續閃電躍升數為前提之決策樹(降雨門檻值為40mm/hr)

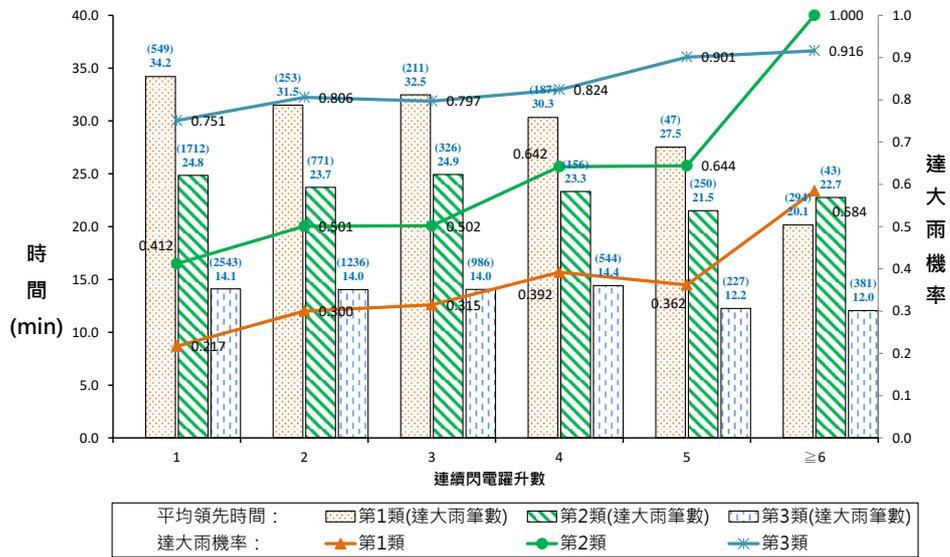


圖5 連續閃電躍升數之決策樹各分類達降雨門檻值機率與平均領先時間(降雨門檻值為40mm/hr)

表1 2017年午後對流10分鐘降雨量對應發生大雨之機率

10分鐘降雨量	已發生大雨(含t0)	未來10分鐘	未來20分鐘	未來30分鐘	未來40分鐘	未來50分鐘	未達大雨
≥ 6	0.008	0.022	0.072	0.119	0.149	0.160	0.840
≥ 8	0.015	0.051	0.126	0.182	0.213	0.223	0.777
≥ 10	0.032	0.109	0.213	0.270	0.300	0.309	0.691
≥ 12	0.075	0.209	0.326	0.376	0.402	0.410	0.590
≥ 14	0.140	0.318	0.435	0.475	0.499	0.504	0.496
≥ 16	0.223	0.428	0.516	0.552	0.577	0.580	0.420
≥ 18	0.315	0.507	0.570	0.601	0.629	0.633	0.367
≥ 20	0.410	0.586	0.621	0.652	0.679	0.683	0.317
≥ 22	0.496	0.645	0.666	0.686	0.717	0.722	0.278
≥ 24	0.569	0.673	0.694	0.710	0.731	0.737	0.263
≥ 26	0.604	0.709	0.726	0.743	0.770	0.778	0.222
≥ 28	0.653	0.733	0.750	0.773	0.790	0.795	0.205
≥ 30	0.734	0.797	0.811	0.825	0.832	0.839	0.161
≥ 32	0.822	0.869	0.879	0.879	0.888	0.897	0.103
≥ 34	0.911	0.933	0.933	0.933	0.944	0.944	0.056
≥ 36	0.933	0.947	0.947	0.947	0.960	0.960	0.040
≥ 38	0.984	1.000	1.000	1.000	1.000	1.000	0.000
≥ 40	1.000	1.000	1.000	1.000	1.000	1.000	0.000

表2 連續閃電躍升數等於1之決策樹各分類筆數與達降雨門檻值比例(降雨門檻值為40mm/hr)

連續閃電躍升數=1	筆數	達大雨筆數 (否,是)	達大雨比例 (否,是)
		10066	(5262,4804)
分類1：10分鐘雨量≤4.25mm	2522	(1973,549)	(0.782,0.218)
分類2：4.25mm<10分鐘雨量≤10.95mm	4157	(2445,1712)	(0.588,0.412)
分類3：10分鐘雨量>10.95mm	3387	(844,2543)	(0.249,0.751)

# Using Data Mining and Decision Tree for Convective Afternoon Rainfall Prognosis in Taiwan

Ming-Sheng Yeh<sup>1</sup> Shin-Gan Chen<sup>2</sup> Ping-Yu Lin<sup>2</sup> Chin-Tzu Fong<sup>2</sup> Treng-Shi Huang<sup>2</sup> Cheng-Han Liu<sup>1</sup>  
Anne-Ru Cheng<sup>1</sup>

Manysplendid Infotech, Ltd.<sup>1</sup>

Central Weather Bureau<sup>2</sup>

## Abstract

In our study, we count total lightning and find out the maximum 10-minute rainfall at the 0.05° resolution grid center with 5km radius in the afternoon convective events in 2017. Then, used the algorithm by Gatlin (2006) to calculate lightning jump. Decision tree methodology is a commonly used data mining method for establishing a classification system and prediction. For improve the technology of the early warning of the short-duration heavy rainfall in Taiwan, we used decision tree as a tool to analyze the different factor combinations of total lightning, lightning jump, and rainfall observation by base on the heavy rain criteria (40 mm/hr), and evaluated the effect of the combination of lightning jump and other factors.

Base on the probability analysis of 10-minute rainfall data, if the recent observed 10-minute rainfall  $\geq 10$  mm, the probability of heavy rain occurring in the future 50 minutes exceeds 31%. If lightning jump occurred, the probability of heavy rain occurring in the future 50 minutes exceeds 50%. Considering the 10-minute rainfall and lightning jump simultaneously, the decision tree improve the confidence of the early warning of heavy rain. If the recent observed 10-minute rainfall  $\geq 10$  mm and lightning jump simultaneously, the probability of heavy rain occurring in the future 60 minutes exceeds 70%. This study demonstrated the effect of applied the 10-minute rainfall data and lightning jump for the early warning of the short-duration heavy rainfall in Taiwan.

Keywords: decision tree, lightning jump, convective afternoon rainfall