

# 以卷積神經網路辨識都市區 CCTV 影像積淹水事件

Using Convolution Neural Network for flooding detection in CCTV images in urban areas

國立臺灣海洋大學  
河海工程學系  
助理教授

**蘇元風\***  
Yaun-Fong Su

國立臺灣海洋大學  
地理資訊系統研究中心  
資深工程師

**簡大鈞**  
Ta-Chun Chien

國立臺灣海洋大學  
地理資訊系統研究中心  
專案經理

**林怡廷**  
Yi-Ting Lin

## 摘要

卷積神經網路 (Convolution Neural Network, CNN) 在許多領域已有相當多成功的應用，在防災領域尚有許多可能的應用面向。臺灣在面對防災工作時，時常仰賴 CCTV 畫面提供即時的現場資訊，然而全台各單位廣泛設立 CCTV 監控系統，目前已有數萬支 CCTV 在監看各地的狀況，然而如此龐大的數量在災害發生當下，仍僅能仰賴人工挑選已知發生災害事件 (例如：積淹水) 地點鄰近的 CCTV 監視畫面，嘗試找到剛好拍攝到積淹水的現場資訊。在災害時的應用上仍較為被動，為解決此問題，本研究借助卷積神經網路對於影像辨識的能力，蒐集 2023 年六場颱風事件以及三次豪雨事件之 CCTV 畫面共計 52,941 張影像，並從中篩選 3,540 張都市區 CCTV 影像進行 CNN 模式訓練，此模式約有 240 萬個參數，訓練與測試結果顯示整體辨識都市區積淹水的準確度可達 98%，針對積淹水事件的正確辨識率為 85%，未來將朝向兩階段 CNN 模式的方式，期能改善此辨識率，使 CCTV 監測各地是否發生淹水事件能更為自動化，增進防災監測效率。

關鍵詞：卷積神經網路、CCTV、都會區積淹水。

\* 通訊作者，國立臺灣海洋大學河海工程學系助理教授  
202301 基隆市中正區北寧路 2 號，yuanfongsu@mail.ntou.edu.tw

# Using Convolution Neural Network for flooding detection in CCTV images in urban areas

**Yaun-Fong Su\***

National Taiwan Ocean  
University  
Department of Harbor and River  
Engineering

**Ta-Chun Chien**

National Taiwan Ocean  
University  
Geographic Information System  
Research Center

**Yi-Ting Lin**

National Taiwan Ocean  
University  
Geographic Information System  
Research Center

## ABSTRACT

Convolution neural network (CNN) has been proved to be very useful in many research fields. However, there are still many possible applications in disaster monitoring and emergency operations. In Taiwan, CCTV also plays an important role for providing real-time situation in the field. Due to its importance, there are more than 20,000 CCTV working in Taiwan to provide real-time information. During flooding events, however, operation agency still needs manpower to check every CCTV footage and it is time and labor consuming. To address this issue, we collected 52,941 CCTV images during typhoon and storm events in 2023. From these images, we selected 3,540 images within urban area during daytime as the training dataset of the CNN model for flooding detection in CCTV image in urban areas. The developed CNN model has 2.4 million parameters to train and the training time costs about 1 hour. The trained CNN model shows a good performance with overall accuracy reaching 98 % while the accuracy for flooding detection for training and testing datasets reaches 85 %. This model may be further improved by considering two-stages CNN model and hopefully may make flooding detection using CCTV in a more efficient way.

Keywords: Convolution neural network, CCTV, Flooding in urban areas.

Su, Y.F.\*, Chien, T.C., & Lin, Y.T. (2024). "Using Convolution Neural Network for flooding detection in CCTV images in urban areas" *Journal of Taiwan Agricultural Engineering*, 70(2), 30-39.

[https://doi.org/10.29974/JTAE.202406\\_70\(2\).0003](https://doi.org/10.29974/JTAE.202406_70(2).0003)

## 一、緒論

隨著氣候不停的在改變，臺灣每年持續受到颱風及豪雨事件影響造成積淹水災情，這些積淹水事件普遍發生於都市地區，由許多國內外的淹水災害事件也顯示，都市區面對淹水災害首當其衝的都是都市地區的交通骨幹。都市交通是維持都市脈動相當重要的骨架，日常生活、農業與工業發展等經濟活動都與交通脫離不了關係。

都市交通的積淹水監測在過去已有許多資源的投入，例如：淹水感測器、下水道水位監測系統等，然而這些儀器尚未普及到都會區的各地，仍集中於過去經驗可知較常積淹水的地區。臺灣目前在都會區有許多正在運行的閉路電視監測系統 (CCTV)，許多單位都有安裝架設 CCTV 監視系統，例如：公路局、地方政府、水利署、水保署、警政署等，為了不同用途而架設 CCTV，全臺至少有 2 萬支 CCTV 隨時在提供現場第一手資訊。其中，公路局的 CCTV 影像也開放免費取用 (<https://data.gov.tw/dataset/29817>)，對於提供都市地區道路積淹水的現場第一手資訊而言，是非常重要的資料來源之一。通常在颱風豪雨期間，防災相關單位會時常緊盯各地的 CCTV 畫面，隨時掌握現場的即時資訊，然而，隨著 CCTV 數量越來越多，以人力監測的做法不僅耗時且需消耗大量人力，更可能偶有疏漏掉積淹水畫面，若能達到自動監測的效果，將可大幅提升防災的即時性與全面性。

隨著深度學習網路技術的發展，卷積神經網路 (Convolutional Neural Network, CNN) 在近年大量地出現在各種領域的應用，都有不少成功的案例，例如：李哲緯(2016)應用於手語影像辨識、Castelluccio 等人 (2015) 與 Xu 等人 (2018) 均應用 CNN 於衛星影像之

土地利用分類、Sarker 等人 (2019)、Wang 等人 (2021)、Lv 等人 (2022) 使用 CNN 為基礎的架構於雷達影像中辨識大面積的淹水事件。此外，林旭信等人 (2023) 則應用深度學習網路於氣象資料，推估西北太平洋颱風強度。CNN 最大的優點在於架構可彈性變化與修改，對於不同的應用目的有高度的適用性，例如：Lv 等人 (2022) 提出修正的 CNN 模式 (Modified DeepLabv3+) 發現對於淹水偵測有良好的成果。此外，以 CNN 為基礎的知名深度學習網路 U-Net，也在影像分割領域有許多的應用，例如：Zhou 等人 (2023) 比較多種以 CNN 為基礎的深度學習網路對於衛星影像分類的表現。有關 CCTV 與 CNN 結合的應用則有 Singh 等人於 2020 年提出可應用於偵測 CCTV 畫面中的異常偵測，以及廖昱唐 (2021) 以工地攝影機結合以 CNN 為基礎的 YOLO 深度學習網路，辨識未配戴工地安全帽的人員。

為解決上述問題，本研究蒐集 2023 年共 6 場颱風以及 3 場豪雨事件共計 52,941 幅 CCTV 影像，並從中篩選出都市區且於日間拍攝的影像，共計有 3,540 張影像，其中包含 205 張積淹水影像，以及 3,335 張非積淹水影像。這 3,540 張影像將作為卷積神經網路的訓練與測試資料，建置自動辨識 CCTV 監視畫面的積淹水事件模型。

## 二、研究區域與材料

### 2.1 研究區域

本研究選擇以北部地區의 CCTV 為主，共挑選了 119 個 CCTV 測站，其分布如圖 1 所示，其中新北市計有 69 個測站、桃園市 43 個、基隆市則為 7 個，各

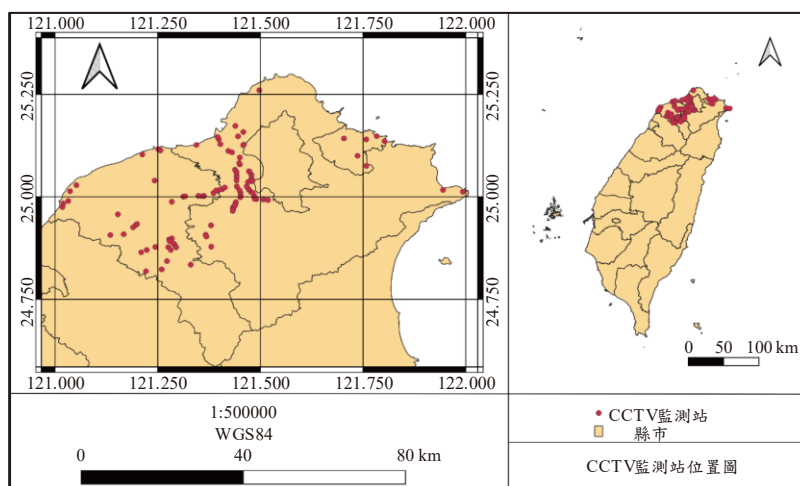


圖 1 研究區域與 CCTV 監測站位置

表 1 本研究使用之各縣市 CCTV 測站

縣市	區 (數量)	總計
新北市	八里區(4)、三芝區(1)、三重區(4)、三峽區(4)、土城區(13)、中和區(7)、五股區(6)、板橋區(13)、泰山區(2)、貢寮區(2)、林口區(1)、淡水區(4)、新莊區(8)	69
桃園市	大園區(1)、大溪區(14)、平鎮區(3)、桃園區(3)、新屋區(5)、楊梅區(2)、龍潭區(4)、龜山區(7)、蘆竹區(3)、觀音區(1)	43
基隆市	中正區(3)、安樂區(1)、暖暖區(2)	7



圖 2 CCTV 拍攝之積淹水畫面 (A-B) 市地區 (C-D) 山區



圖 3 CCTV 畫面夜間影像範例

區所挑選的 CCTV 數量則列於各區後面括號內，詳細數據請見表 1。CCTV 測站挑選時選擇在 2023 年的颱風豪雨事件中經常被記錄到 CCTV 測站為主。

## 2.2 訓練影像篩選

回顧 2023 年，臺灣共歷經了六次颱風事件（分別是瑪娃颱風、杜蘇芮颱風、卡努颱風、蘇拉颱風、海葵颱風以及 10 月的小犬颱風）以及三次豪雨事件（0612 豪雨、0907 豪雨、0910 豪雨），期間共蒐整了全

台共計 52,941 張 CCTV 影像，這些影像包含山區與都市區，山區與都市區的影像對比如圖所示，可以發現除了畫面中的街景截然不同，當積淹水發生時，兩者積淹水型態也不同，例如：山區積淹水的水體較為混濁、都市區的積淹水則普遍較為清澈，且時常有機車或汽車通過造成的水花與漣漪，這些影像特徵都將可能被卷積神經網路學習而作為判斷積淹水的依據，因此針對都市區的積淹水研究而言，本研究僅使用都市區內的 CCTV 鏡頭進行後續研究。此外，由於夜間影像的照明條件不佳（請見圖 3），且容易受到車燈與路

燈的影響，因此在本研究中也並不納入夜間影像。最後，本研究篩選出都市區且於日間拍攝的影像，共計有 3,540 張影像，其中包含 205 張積淹水影像，以及 3,335 張非積淹水影像，將做為後續 CNN 模式建置之用。

### 三、研究方法

#### 3.1 卷積神經網路

卷積神經網路 (Convolution neural network, CNN) 是建立在類神經網路與影像分析這兩大領域之上所發展出來的深度學習網路架構。最早提出類似 CNN 架構的文章是由 Fukushima (1980) 提出的 S-cell 與 C-cell 架構，其中 S-cell 即為 CNN 架構中的卷積層 (Convolution layer) 而 C-cell 即為 CNN 中的池化層 (pooling layer)。CNN 一詞最早由 LeCun 等人於 1998 年所提出的 LeNet-5 架構中出現，其透過輸入一張影像後，在卷積層中萃取出多種影像資訊，並透過池化層

將重要資訊保留且減少影像大小，達到加速運算的目的，經由多組的卷積層與池化層組合，萃取出影像中多層次的特徵，這些特徵將被整理為傳統類神經網路的輸入資料，而最後的輸出即為類神經網路的輸出層所定義的輸出值，其架構如圖 4 所示。

卷積層的操作是將影像經過許多不同的空間濾波器 (也稱作核心) 進行卷積 (convolution) 計算，一般而言空間濾波器的大小通常選擇為 3 乘 3 的大小，卷積計算可以表示如下

$$V' = \sum_{i=1}^9 w_i x_i \dots\dots\dots (1)$$

其中， $w_i$  為空間濾波器 (有時也稱為 Kernel) 內的數值，此數值普遍介於 -1 ~ 1 之間的實數， $x_i$  則為影像中的灰階值 (digital number or grey level)，其計算過程為自影像的左上角開始，找到第一個 3 乘 3 的影像矩陣，這 9 個灰階值將與核心的 3 乘 3 矩陣，依據對應的位置相乘後，再加總而得特徵值 ( $V'$ )，如圖 5 所示。每個核心的 3 乘 3 矩陣可由 -1 至 1 的任意數字組

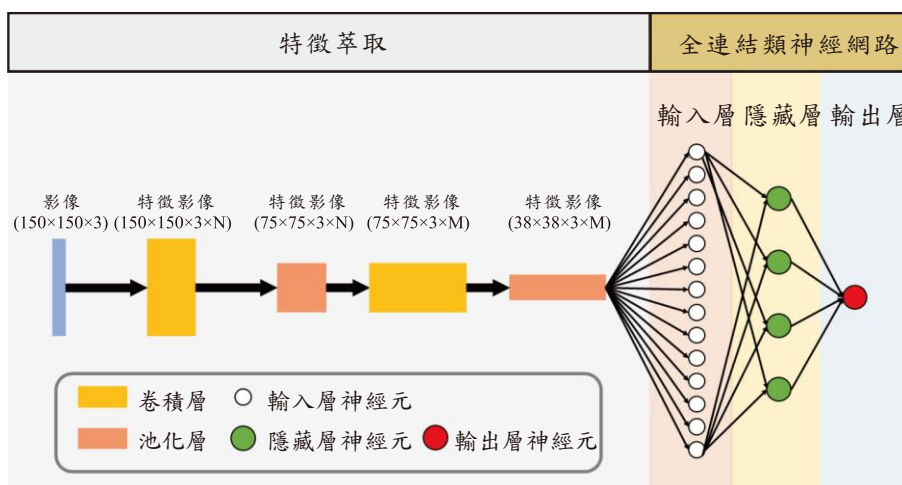


圖 4 CNN 模式架構示意圖

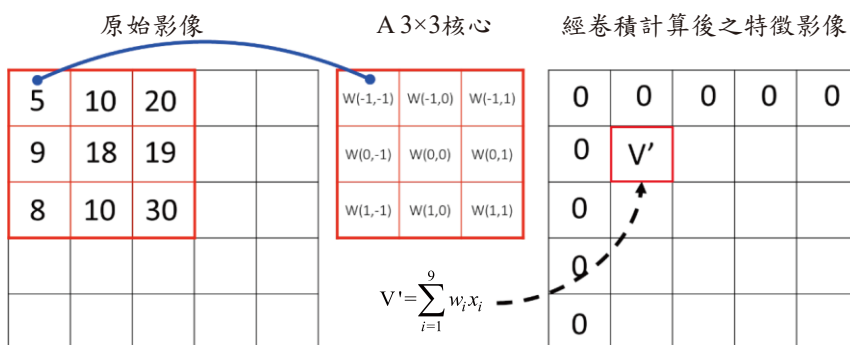


圖 5 影像卷積計算示意圖

合而成，因此理論上可以有無限多種組合，每種核心的組合可用於萃取出不一樣的影像特徵，例如：若 3 乘 3 矩陣核心內的數值都是 0.111 時，則此時得到的特徵是影像的平均值。

池化層的計算則是選擇 2 乘 2 的矩陣，將四個數字取最大值，因此也稱為 **Max Pooling**，即最大值的池化層。

$$P_i = \max \{v_i\}, \quad i = 1, 2, 3, 4 \dots\dots\dots (2)$$

此計算的重點在於保留最顯著的資訊，並且同時將影像大小縮小為原本的四分之一，故可達到加速計算速度的效果。其計算過程如圖 6 所示，以一個 6 乘 2 的影像為例，首先取出 4 個灰階值，並保留最大值，而後往下取出另一組 2 乘 2 的影像灰階值，並保留最大值，最後一個 6 乘 2 影像矩陣經過池化計算後為一個 3 乘 1 的矩陣。

CNN 模式訓練過程中，必須指定損失函數 (loss function) 以及精度 (accuracy) 指標，以利 CNN 模式之參數調整有所依據 (Lin 等人, 2018)，由於本研究主要目的為辨別影像是否為淹水，屬於二元分類問題，因此採用的損失函數為二元交錯熵 (Binary Cross Entropy, CE)，精度則是總體準確度 (overall accuracy, Accu)，可表示如公式 (3)：

$$CE(p, y) = \begin{cases} -\log(p), & \text{if } y = 1 \\ -\log(1-p), & \text{otherwise} \end{cases} \dots\dots\dots (3)$$

$$Accu = \frac{N(y = y_{pred})}{N_{total}} \dots\dots\dots (4)$$

其中， $p$  為模式輸出的屬於該類別之機率， $y$  則是真值 (ground truth)， $y_{pred}$  則是模式輸出之類別， $N(y = y_{pred})$  則是輸出值與真值相符得資料個數， $N_{total}$  為訓練之總樣本數。

### 3.2 準確度評估與門檻值挑選

準確度評估方法採用混淆矩陣 (Confusion matrix)，如表 2 所示，其中 TP 與 TN 分別為 True Positive 與 True Negative，屬於正確分類，FP 與 FN 分別為 False Positive 與 False Negative，屬於錯誤分類，其中可定義使用者準確度 (user accuracy)、生產者準確度 (producer accuracy) 以及總體準確度 (overall accuracy)，將用於評估 CCTV 辨識淹水時之準確度。

在 CNN 模式中，輸出值為 0~1 之間的數值，其可視為 CNN 辨識淹水與否的機率值，然而此機率值要轉換為是否為淹水時需要一個門檻值，將輸出結果

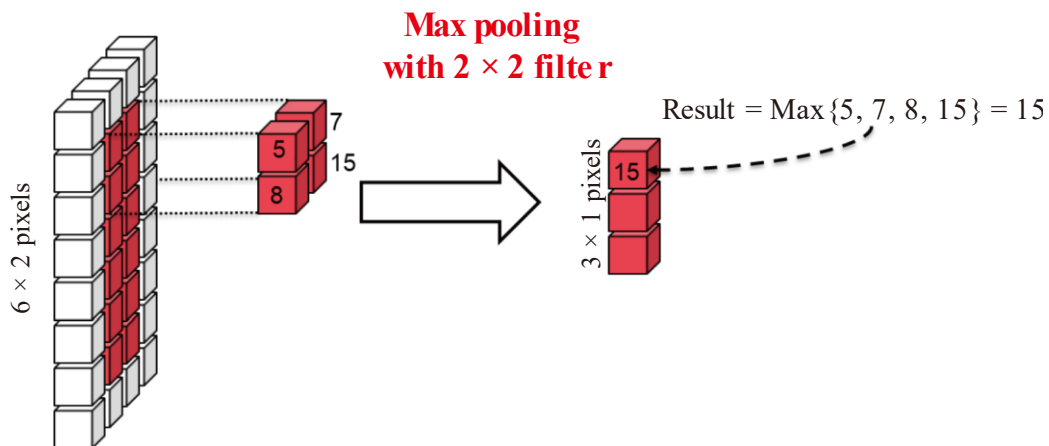


圖 6 影像池化計算示意圖

表 2 二元分類問題之混淆矩陣

		實際情形			
		非積淹水	積淹水	總計	使用者準確度
分類	非積淹水	TP	FP	TP+FP	TP/(TP+FP)
	積淹水	FN	TN	FN+TN	TN/(FN+TN)
	總計	TP+FN	FP+TN	N	
	生產者準確度	TP/(TP+FN)	TN/(FP+TN)	整體準確度 = (TP+TN)/N	

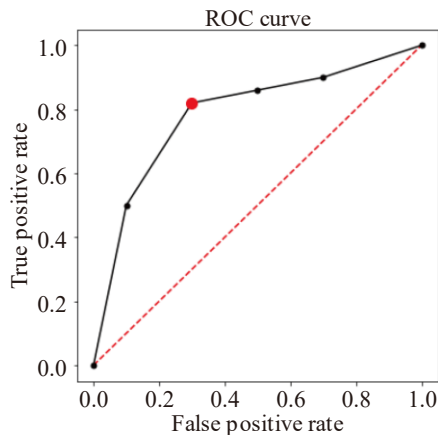


圖 7 ROC 曲線示意圖

二值化，其門檻值的選擇方式可透過接收者操作特性曲線 (Receiver Operator Characteristic Curve, ROC) 進行選擇。ROC 曲線如圖 7 所示，其橫軸為偽陽性 (False Positive Rate, FPR)、縱軸為真陽性 (True Positive Rate, TPR)，而最佳的門檻值即為 ROC 曲線中最靠近左上角點時的門檻值，值得一提的是為了使門檻值較為客觀並減少過度擬合 (overfit) 的問題，ROC 曲線是使用測試資料進行計算。

#### 四、結果與討論

卷積神經網路的建置與訓練工作採用 Python 的 Keras 套件進行分析，由於 CCTV 影像大小不一，為利於後續計算，本研究統一調整影像大小至 150 乘 150 的大小，影像數量共有 3,540 張，包含 205 張積淹水畫面以及 3,335 張非積淹水畫面，CNN 模式建置時採用 80 % 的資料作為訓練，20 % 的資料做為測試。CNN 模式架構如圖 8。

模式中包含三組的卷積層與池化層處理，接著以全連結神經網路，具有 128 個神經元的隱藏層進行分析，最終以 Sigmoid 函數將模式輸出值限縮於 0~1 之間的數值，用於判斷是否為積淹水的門檻值，經過 ROC 曲線分析後，門檻值採用 0.35，若模式輸出值大於 0.35 將判斷為積淹水，反之則為非積淹水。為了避免過度擬合 (overfitting) 的問題，模式架構過程採用 dropout 的參數，其用意在於隨機捨棄 20 % 或 40 % 的權重連結，以減少過度擬合發生的情形。模式建置過程中採用 Adam 學習法，模式中共計有 2,424,065 個參數需要訓練，總計訓練時間約為 1 小時，使用一般桌上型電腦，CPU 為 Intel® Core™ i5-10500，記憶體為 8 GB。

模式訓練過程之誤差與準確度隨著疊代次數 (Epochs) 的變化如圖所示，其誤差的變化趨勢為逐漸降低，而準確度也是逐漸增加且訓練與驗證的趨勢一致，由此可見模式的訓練過程並沒有觀察到過度擬合的問題，顯示此模式是穩定的。訓練與測試之分類準確度則如表 3 所示。由表 3 可知，模式在訓練時幾乎可以將訓練用之 2,832 張影像正確地進行分類，只有兩張非積淹水影像被錯誤分類為積淹水的影像，當此訓練好的模式應用於測試資料時，41 張的淹水影像有 5 張被錯誤分類為非淹水 (約佔 12 %)，而在 667 張非淹水影像中則僅有 6 張錯誤判斷為積淹水 (約佔 1 %)，反過來說，在 708 張測試影像中有 42 張被分類為積淹水，但是其中有 6 張是錯誤分類的 (約佔 15 %)，顯示此模式仍有進步空間。

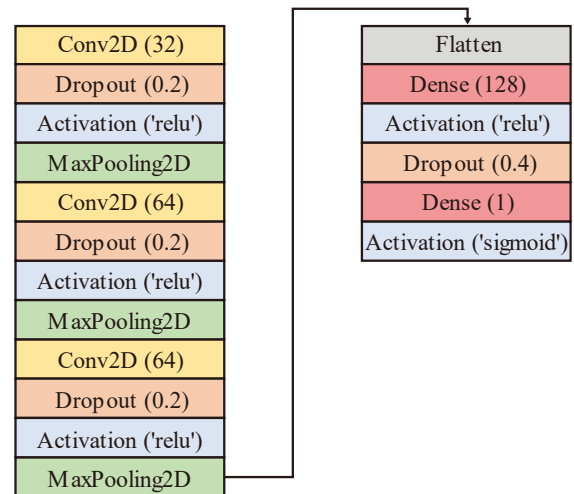


圖 8 以 Keras 建置之 CNN 模式架構

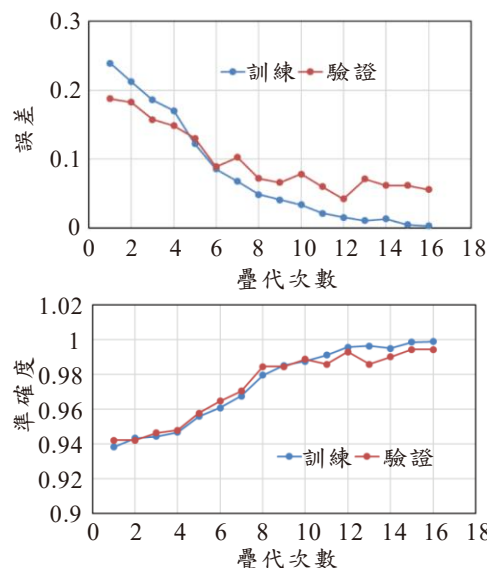


圖 9 訓練與驗證資料之訓練歷程

表 3 CNN-Flooding 模式訓練與測試資料之分類準確度評估

訓練		實際情形			
		非積淹水	積淹水	總計	使用者準確度
分類	非積淹水	2666	0	2666	100 %
	積淹水	2	164	166	98.80 %
	總計	2668	164	2832	
	生產者準確度	99.93 %	100%	總體準確度 = 99.93 %	
測試		實際情形			

		非積淹水	積淹水	總計	使用者準確度
分類	非積淹水	661	5	666	99.25 %
	積淹水	6	36	42	85.71 %
	總計	667	41	708	
	生產者準確度	99.10 %	87.80 %	總體準確度 = 98.45 %	

在測試資料集中被錯誤分類的 11 張影像



圖所示，可以發現被錯誤分類的影像中，有些可能因為雨

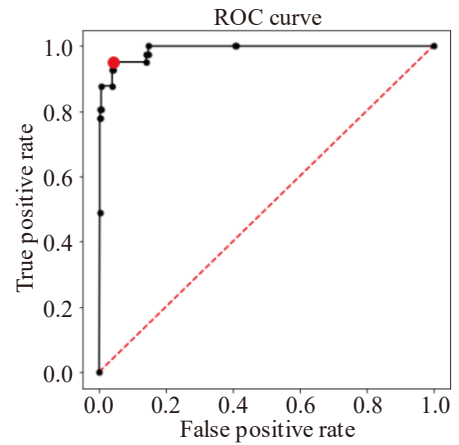


圖 10 ROC 曲線結果

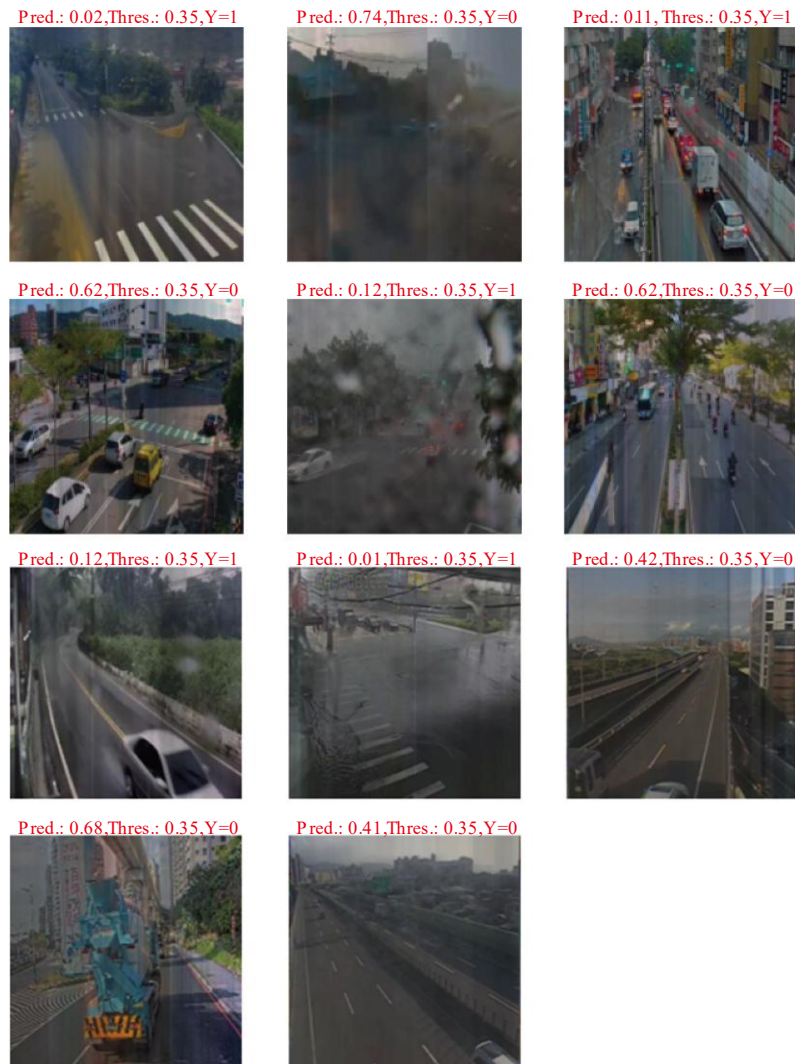


圖 11 淹水 CNN 模式於測試資料集中被錯誤分類的 11 張影像，其中 Pred 為模式輸出值，Thres 為判斷門檻值，Y 為影像標註值 (1 代表積淹水、0 代表非積淹水)

滴遮蔽鏡頭，有些則是都市積淹水時的水色較為混濁，有些則是影像拍攝時天色較為昏暗，車燈與路面反射造成誤判。整體而言，整體準確度雖然可達 98%，但是在積淹水影像的錯誤分類的情形仍有 15% 左右的誤判，未來仍有持續改善的地方。若與前人的類似研究比較，本研究的準確度仍有略高於 Hakim and Dewi (2021) 使用 CNN 用於偵測 CCTV 影像中是否下雨的模式，他們的模式在訓練時準確度為 98.8%，測試資料則為 96.7%，整體則是 96.4% 的準確度，而本研究不僅使用較多影像且用於偵測較為複雜的淹水現象，整體 (包含訓練與驗證) 仍有 99.6% 的準確度。

## 五、結論

本研究使用卷積神經網路 (CNN) 於 CCTV 影像，建置都市區積淹水事件的自動化判釋模式，研究中採用 3,540 張都市區 CCTV 影像用於建置 CNN 模式，由模式的建置過程與分析成果，彙整以下重點結論：

1. 建置 CNN 模式於 CCTV 畫面積淹水判斷時，由於 CCTV 畫面拍攝時的特徵容易因為山區或是都市區而有差異，此外，日間與夜間拍攝的影像差異較大，因此建置 CNN 模式時，應適當挑選訓練與測試之影像。卷積神經網路具有高度的應用潛力，可有效的監測多支 CCTV，快速地判斷 CCTV 畫面中是否出現路面積淹水的情形，測試時誤判率約為 15%，仍有改善空間。對於非積淹水影像的判斷上，整體而言 CNN 模式的精度可達 98%，但針對積淹水影像仍有 15% 左右的誤判率，未來改善方向將朝著採用兩階段 CNN 模式

的方式進行，在第一階段 CNN 模式將用於辨識是否為一般晴朗的天氣或是有下雨的情形，第二階段再進行畫面中是否為積淹水，透過兩階段的判斷期能提高積淹水的辨識精度。

## 誌謝

本研究承蒙國土管理署計畫（計畫編號：111-4000-0402-0000-0010）之經費補助，得以完成特此致謝。

## 參考文獻

1. Castelluccio, M., Poggi, G., Sansone, C., Verdoliva, L., “Land Use Classification in Remote Sensing Images by Convolutional Neural Networks,” *arXiv preprint arXiv:1508.00092*, 2015.
2. Fukushima, K., “Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position,” *Biological Cybernetics*, Vol. 36, pp. 192-202, 1980.
3. Hakim, A.L., Dewi, R., “Automatic Rain Detection System Based on Digital Images of CCTV cameras Using Convolutional Neural Network Method,” *IOP Conf. Series: Earth and Environmental Science* 893, 012048, 2021.
4. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., “Gradient-based Learning Applied to Document Recognition,” *Proceedings of the IEEE*, Vol. 86, no. 11, pp. 2278-2324, 1998.
5. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollar, P., “Focal Loss for Dense Object Detection,” *arXiv:1708.02002v2*, 2018.
6. Lv, S., Meng, L., Edwing, D., Xue, S., Geng, X., Yan, X., “High-performance Segmentation for Flooding Mapping of HISEA-1 SAR Remote Sensing Images,” *Remote Sensing*, Vol. 14, pp. 5504, 2022.
7. Sarker, C., Mejias, L., Maire, F., Woodley, A., “Flood Mapping with Convolutional Neural Networks Using Spatio-contextual Pixel Information,” *Remote Sensing*, Vol. 11, pp. 2331, 2019.
8. Singh, V., Singh, S., Gupta, P., “Real-time Anomaly Recognition Through CCTV Using Neural Networks,” *Procedia Computer Science*, Vol. 173, pp. 254-263, 2020.
9. Wang, J., Wang, S., Wang, F., Zhou, Y., Wang, Z., Ji, J., Xiong, Y., Zhao, Q., “FWENet: A Deep Convolutional Neural Network for Flood Water Body Extraction Based on SAR Images,” *International Journal of Digital Earth*, Vol. 15, No. 1, pp. 345-361, 2021.
10. Xu, X., Li, W., Ran, Q., Du, Q., Gao, L., Zhang, B., “Multisource Remote Sensing Data Classification Based on Convolutional Neural Network,” *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 56, no. 2, pp. 937-949, 2018.
11. Zhou, H., Yang, J., Zhang, T., Dai, A., Wu, C., “EAS-CNN: Automatic Design of convolution AI Neural Network for Remote Sensing Images Semantic Segmentation,” *International Journal of Remote Sensing*, Vol. 44, No. 13, pp. 3911-3938, 2023.
12. 李哲緯，「基於卷積神經網路之手語影像辨識」，國立成功大學碩士論文，2016。
13. 林旭信、溫慶亞、朱楷洋、何御東，「以深度學習神經網路推估西北太平洋颱風強度」，*農業工程學報*，第 69 卷，第 1 期，第 29-40 頁，2023。
14. 開放資料平台，<https://data.gov.tw/dataset/29817>，最近瀏覽日期：2024 年 4 月 10 日。

15. 廖昱唐，「基於卷積神經網路及影像強化之工地安全帽偵測系統及應用」，國立臺灣大學機械工程學研究所碩士論文，2021。