



National Chiao Tung University

Department of Transportation & Logistics Management

Working Paper

## Using convolutional neural network to construct classification models of bus drivers' car-following behavior

Wen-Chin Lin and Chung-Cheng Lu

Working paper NCTU-TLM-WP-19-021

July 15 2019

## 摘要

國道客運安全為我國所重視之交通議題之一，若能了解國道客運駕駛之駕駛情況，客運業者便能夠有效對駕駛員採行適當的管理策略。本研究之目的在以卷積神經網路(CNN)建構一套跟車行為分類模型，將車載系統 Mobileye 所記錄之駕駛員跟車車距資料作為研究資料，並考慮可能影響跟車行為之因素，如：天氣、車速及日夜三個因素，進一步設計七種不同的情境組合來篩選資料並產製跟車車距矩陣圖，將跟車車距矩陣圖輸入至本研究訓練之 CNN 進行分類模型，得出該駕駛員的跟車類型。

經本研究實驗，本研究所提出之分類模式在大多情境下皆有不錯的分類成果，模型分類精確率平均為 97%，顯示本研究之分類模型具備駕駛員跟車行為分類之能力，此模型可提供客運業者作為參考，藉由分類結果評估駕駛員的跟車行為類別，並針對不同類別駕駛員採行適當的管理策略。

**關鍵字：**機器學習、資料探勘、卷積神經網路、交通安全、公共運輸

## 壹、緒論

### 一、研究背景與動機

近年來政府積極推動公共運輸，鼓勵民眾搭乘公共運具，減少私人運具之使用量以達到節能減碳效果。臺灣的公共運輸系統，在政府的鼓勵及推動之下日益健全，根據民眾日常使用運具狀況調查摘要分析（105 年）指出臺灣公共運輸市占率達 15.9%。目前國內城際間的公共運輸主要仰賴高鐵、臺鐵及汽車客運，根據交通部統計月報（107 年）指出，107 年之汽車客運量達 1,244,630 千人次，高於臺鐵的 231,268 千人次及高鐵的 63,963 千人次，由此可知民眾使用汽車客運之頻率較其他公共運輸運具高出許多。

臺灣汽車客運業中，以國道客運肩負起臺灣城際運輸的重要角色，但近年來汽車客運發生事故造成重大傷亡的案件層出不窮，根據交通部統計月報（107 年）顯示，近五年大客車平均每年肇事件數為 27.5 件，平均死亡人數達 35 人，平均每萬輛死亡率為 7.16%；而小客車平均肇事件數為 413 件，平均每萬輛死亡率為 2.04%，單以每萬輛死亡率來比較，發現近五年之大客車之平均每萬輛死亡率遠高於小客車，顯示大客車交通事故對社會影響巨大，近五年大客車交通事故概況，如表 1 所示。

表 1 近五年大客車道路交通事故發生概況

| 統計年份 | 肇事件數 | 每萬輛肇事率 | 死亡人數 | 每萬輛死亡率 | 受傷人數 | 每萬輛受傷率 |
|------|------|--------|------|--------|------|--------|
| 107  | 22   | 6.46   | 22   | 6.46   | 1    | 0.29   |
| 106  | 28   | 8.15   | 67   | 19.5   | 53   | 15.43  |
| 105  | 33   | 1.45   | 33   | 1.45   | 29   | 0      |
| 104  | 24   | 7.18   | 25   | 7.48   | 29   | 8.68   |
| 103  | 29   | 0.91   | 29   | 0.91   | 3    | 0      |

根據道路交通事故肇事原因與肇事者特性分析（106 年），顯示 A1 類肇事原因高達 94% 為駕駛人過失所致；而據世界衛生組織報告（2011 年），指出駕駛在開車期間約有 30% 的時間分心在做其他事；美國總務管理局（GSA）提出了六項不安全的駕駛行為，包括不適當的車速、違反道路優先權、偏離車道中心、轉彎不當、跟車過近、不當行駛等。在國內因「跟車過近」而發生的交通事故，往往造成的嚴重傷亡，106 年 9 月 11 日深夜國道車禍造成 6 死 11 傷之慘劇，肇因為駕駛員分心未保持與前車距離釀禍；107 年 5 月，國道一號發生客運車與三輛大貨車之追撞事故，造成九人受傷，疑似因未保持安全跟車車距離而釀禍，由上述案例可知駕駛員的駕駛情況對於民眾乘車之安全影響巨大，也是目前政府和業者急需面對的迫切問題，確實有必要針對可能影響駕駛員行車之因子進行分析。

此外因應 AI 世代的來臨，人工智慧（AI）已能建構與人類相似甚至超越人類的推理、學習、辨識及預測等能力，曾經人類對人工智慧充滿各種天馬行空的想像，如今皆以不同型式落實在生活中，例如：Google Maps 利用深度學習辨識路牌及門牌，使其能從街景影像中找出地址；進出停車場的車牌辨識系統等，皆是利用人工智慧來達到。隨著人工智慧在各領域應用逐漸受到重視，其中的深度學習方法為人工智慧領域中非常重要且效果良好的工具，近年逐漸應用於交通運輸領域各方面研究，例如：以深度學習 CNN 模式來預測號誌化道路的速率（Ma, Dai, He, Ma & Wang, 2015）；以 CNN 預測市區道路路況之研究（黃冠維, 2018）；以深度學習方法預測短期流量，並加入天氣預報作為變數測試（Polson & Sokolov, 2017），皆有良好的成果。目前深度學習結合交通領域之相關研究大多應用於預測未來路況環境等情形，這些研究都有不錯之成果且足以提供交通主管機關或相關單位作為營運的參考。

國內、外有許多針對駕駛行為之相關研究，但多數為針對駕駛員進行心理量測問卷或針對駕駛員所涉及之事故進一步對潛在危險因子進行分析，例如：評估危險因子對大客車駕駛員發生交通事故的嚴重程度，探討不同危險因子對不同類駕駛員的事故影響程度（Feng, Li, Ci & Zhang, 2016）。有多篇以行車記錄器資料為基礎分析駕駛行為的研究，例如：以行車記錄器結合駕駛員人因特性，建構模式探索其駕駛績效，未保持安全距離，駕駛績效越差（陳信丞, 2014）；以行車記錄器資料為基礎建立駕駛風險分級制度，將偏差行為做集群分析時分為兩群有較佳的效果（李佩憶, 2017）；此外，外在因子也會影響駕駛情形，例如：夜間及車速對事故嚴重程度具影響（Feng et al., 2016）。

過往較少有研究著重在結合機器學習方法於駕駛員行為之分類。因此綜合上述研究背景與動機，本研究擬以深度學習方法建立資料探勘模式，用以分析客運公司車載系統 Mobileye 所記錄與前車距離之跟車資料，建立不同情境下駕駛員跟車行為類型之分類模式，並提供客運業者參考，客運業者可進一步對駕駛員跟車行為類別進行分析，由分類結果得知駕駛員的跟車行為類型，針對高風險的駕駛員採行適當的管理策略，使得客運業者在管理層面能有效地對症下藥。

## 二、研究目的

本研究之目的在於利用 CNN 與客運業者提供之 Mobileye 跟車車距資料來建構一套跟車行為類型分類模型，利用 Mobileye 記錄之資料匯入模型，分析駕駛員之跟車車距，並探討將駕駛員跟車類別分為積極型及保守型兩類，與積極型、穩健型及保守型三類之分類效果，此模式之預測結果可提供客運公司業者了解駕駛員之跟車行為類型所屬類別，並提供客運業者參考，針對不同類別的駕駛員採行適當的管理策略。本研究之目的如下：

- (一) 藉由文獻回顧了解可能影響駕駛員駕駛行為之相關因子。
- (二) 利用集群分析方法將駕駛員跟車車距進行分類，並以 CNN 模型來建構跟車行為類別之分類模型，建構一套有效分類駕駛員跟車類型之分類模型。

## 貳、文獻回顧

本研究針對客運駕駛行為之相關研究、機器學習應用於交通運輸之相關研究進行回顧，和研究中會使用到的資料探勘理論方法進行回顧。

### 一、客運駕駛行為相關研究

本研究利用車載系統 Mobileye 探討客運駕駛員之跟車行為類型，本小節針對與客運駕駛行為之相關研究進行回顧，藉由文獻回顧了解哪些可能是影響可運駕駛之駕駛行為的潛在因素，更進一步了解可以藉由那些資料來從事客運駕駛行為之相關研究。

Vujanić et al. (2016) 欲探討心理特質與駕駛員事故風險是否有相關，也提及專家認為人為因素是影響事故發生的重要因子，該研究針對客運公司之駕駛員進行心理特質測驗。該研究探討駕駛員教育程度、是否曾經有事故經驗，以及駕駛員之年齡等因子對發生事故風險之影響，根據研究結果顯示，年輕駕駛員以及教育程度較高之駕駛員，具備較優異的安全駕駛特質。

李佩憶 (2017) 以偏差駕駛行為為基礎來建立駕駛風險分級制度，透過建構多元回歸、二元羅吉斯、次序性羅吉斯模型檢驗人格特質對國道客運駕駛員之駕駛行為與分級結果的影響，該研究之研究對象為國內知名客運公司之 62 位駕駛員，該研究主要探討人格特質及駕駛行為兩面向，以問卷調查來蒐集駕駛員之五

大人格特質資料，駕駛行為資料則是由客運公司提供之行車記錄器資料，其中該研究依據偏差駕駛行為進行階層集群分析，發現分為兩集群是較恰當的分群結果。對駕駛員偏差駕駛行為為集群分析，僅能分為兩集群，其中 77% 為無特別顯著偏差駕駛行為的集群、23% 駕駛員具有相對顯著偏差駕駛行為，此外其研究結果推測車輛行駛的安全距離與駕駛行為存在相關性。

陳信丞 (2014) 透過行車記錄器所蒐集之駕駛員行為資料與駕駛員自身之人因特性探究其駕駛績效，應用結構方程模式來建構模式，該研究針對某國內客運公司之 119 位駕駛員進行研究，該研究之人因評估包含駕駛員自身人因特性及組織管理人因特性；而行車記錄器之資料則是採用 8 項變數（包含嚴重超速、未保持安全距離等）。其研究結果說明駕駛員之人因特性構面會對駕駛行為產生顯著且正向影響，意即對駕駛日常的違規或危險駕駛行為造成正面影響。駕駛行為構面對駕駛績效產生顯著的負面影響。

## 二、資料探勘理論與方法

本研究使用集群分析方法將資料進行初步分類，集群分析之目的是將資料分成若干相異性最大性的群組，而使每個群組內的相似程度最高，吳明隆 (2018) 在 SPSS 操作與應用—多變量分析實務一書中提及：「研究者如果認為觀察值間並非全部同質，在資料探索分析方面，集群分析是一個非常有用的技巧」。而其方法是先計算各觀察值的相似性及相異性，再經由分析方法將這些觀察值進一步分類，經分類過後的觀察值，在同一集群內會具有高度的相似性或同質性，若在不同集群中則會有高度異質性。

本研究以 K 平均法 (K-means clustering) 來對資料進行分群，其運行方式是先將所有觀察體資料分為 K 個集群，可依照研究者指定的集群數量來進行分類，分類步驟如圖 1。

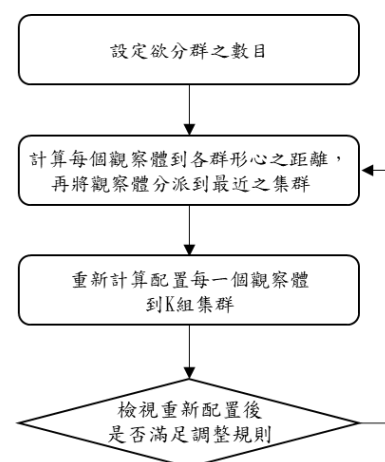


圖 1 K 平均法流程圖

## 三、機器學習方法應用在交通運輸

隨著人工智慧的蓬勃發展，近年因為硬體進步使得機器學習開始受到重視，其中，深度學習就是機器學習的方式之一，也是當今人工智慧的主流，如 AlphaGo 就是最好的例子，透過設定好的神經網路架構，將大量的棋譜資料輸入，讓神經網路進行學習，學習完畢後 AlphaGo 就可以依照不同的棋盤情況作出下一步棋的回應。而機器學習也被廣泛應用在各領域，並且得到良好的成效，如：交通量預測應用，能更有效提供交通主管機關作為決策參考。

Ma et al (2017) 利用 CNN 模式預設號誌化道路速率。將收集得路網數據資料彙整成二維矩陣，並以時空概念呈現此矩陣，可透過矩陣了解各個地點不同時刻的速率資料，再將有速率資訊的矩陣轉換為時空矩陣圖，作為 CNN 之輸入資料。此研究以 CNN 模式，藉由萃取輸入資料的特徵來進行速率預測。

黃冠維 (2018) 利用 CNN 預測市區道路路況，該研究藉由車輛偵測器資料及 CNN 來預測道路未來 5~30 分鐘的速率值。此研究將資料經過清洗處理後，產出速率的時空矩陣，用來呈現同一時間不同空間之速率值、同一空間不同時間之速率值。將產製之時空矩陣轉化為影像概念，產出的時空矩陣圖中速率用顏色深淺來呈現，顏色越深代表速率值越低，顏色越淺代表速率值越高。該研究提出的預測模式大多得到良好的預測結果，平均 MAE 值為 4.93%、平均 MAPE 值為 15.5%，顯示此研究具有準確預測市區道路路況的能力。

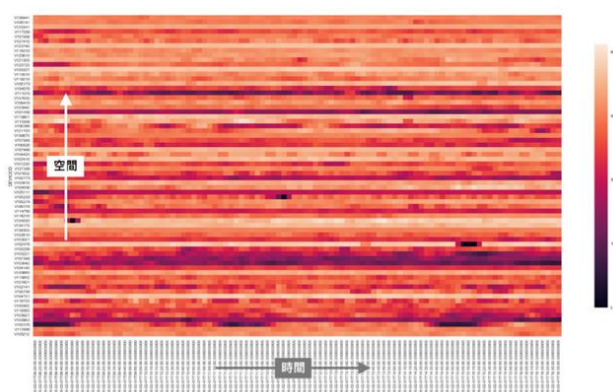


圖 2 時空矩陣圖

Polson et al (2017) 利用深度學習方法預測短期的交通流量，而預測交通流量之研究最大的挑戰為非線性情形，此研究開發一套可捕捉非線性時空效應的深度學習架構，以芝加哥州際公路上的迴線式車輛偵測器所蒐集之資料來進行實驗，並成功將天氣預報作為預測的變量，改善了預測能力，說明此研究所提出的方法能有效達到短期交通流量預測。

以往針對客運駕駛之研究，大多採取心理量測之問卷為主要評估方式，然而隨著科技的進步，各種車載裝置的問世和普及，可利用車載裝置蒐集之大量數據進行相關分析，例如：李佩憶 (2017) 就利用行車記錄器資料進一步將駕駛行為

數據化分析，並使用集群分析方法將駕駛員偏差駕駛行分群，在該研究中也說明了分兩群為較佳的分群數。而本研究擬透過車載裝置 Mobileye 系統所記錄之駕駛員跟車車距資料，其資料特性是記錄了駕駛員當下實際的跟車車距及行車速度等資料，本研究期望透過駕駛員於駕駛期間所記錄之跟車車距資料來做進一步的分析，以期建立有效的分析模式，並讓車載裝置所記錄的大量資料得以充分的分析利用。

綜合以上文獻，國內外已有許多研究客運駕駛員及安全之研究，多應用於駕駛員駕駛行為分析、駕駛員績效評估等研究。而隨著大數據成為顯學，機器學習的方法也被各領域廣泛應用，使用深度學習進行交通分析是前景看好值得發展的方向，目前相關的研究議題投入較少，因此使用深度學習方法來進行駕駛員跟車車距的分類是值得深入研究探討的。

## 參、研究方法

本節說明本研究之研究資料、研究架構、資料處理流程以及 CNN 之模式架構。

### 一、研究資料

本研究使用之跟車車距資料係由國內某知名行駛國道一號台北新竹間之客運公司（簡稱 HT 客運公司），所裝設之車載系統 Mobileye 蒐集得之跟車車距資料，研究用資料為 HT 客運公司所提供之 2019 年 1 月 9 日至 3 月 28 日，共 78 天之 Mobileye 資料，該資料包含駕駛員編號、車號、跟車車距、車速、日期及時間等資訊，共計 16 個欄位資料；也取得該客運公司之班表資料，資料內容包含班表日期、發車時間、路線編號、往返程、駕駛之駕駛員等資料，共計 8 欄位。

而車載系統 Mobileye 為安裝於車輛上之「先進駕駛輔助系統」(Advanced Driver Assistance Systems, ADAS)，結合 30 萬像素的智能攝影機和 EyeWatch 系統，透過電腦視覺演算法及晶元搭載系統，來提供駕駛者車外環境之變化，包含車道偏移警告、前部碰撞警告、車距監測、行人探測及自動巡航控制系統等，並且進一步達到預警功能，其中的車距監測與警告示意圖，如圖 3 所示，當搭載 Mobileye 系統之車輛於行駛中距前車太近，系統會警告駕駛員其車距過近。

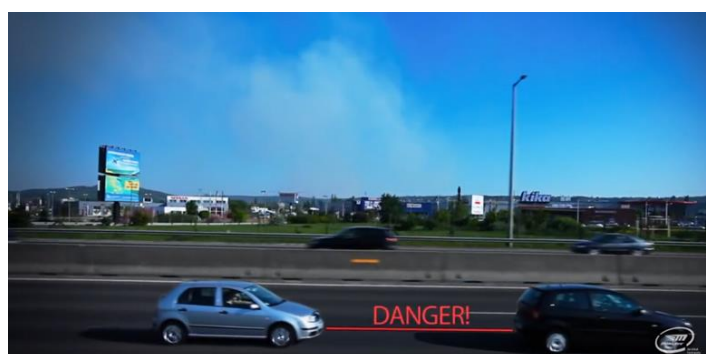


圖 3 車距監測與警告示意圖

本研究選定之研究資料為 Mobileye 系統中之跟車車距資料，其資料蒐集原理是以裝設於車輛上之智能攝影機所拍攝之照片，由電腦視覺演算法來辨識前方物體型態並透過不斷的偵測自身車速及前方車速並透過內部公式的計算來分析前方車況及路況，若與前車距離過近，則會透過車載之 EyeWatch 系統來對駕駛員發出警報，EyeWatch 系統上所顯示的數值 0.8，代表距離前方車輛，還有 0.8 秒的碰撞距離。

## 二、研究架構

本研究目的為藉由車載系統 Mobileye 跟車車距資料以 CNN 模式來建構駕駛員之跟車行為類型分類。研究架構圖如圖 4 所示，詳細說明如下：

(一) 資料蒐集與處理：本研究之輸入資料是利用 HT 客運公司所提供之 2019 年 1 月 9 日至 3 月 28 日，共 78 天之車載系統 Mobileye 跟車車距資料。在處理階段，先針對此跟車車距資料進行清洗，確認資料無異常狀態後，依照本研究設定之情境分類，再產製輸入至 CNN 之跟車車距矩陣圖，並將資料以 K 次交叉驗證方法（K-fold cross-validation）分為 K 份，其中 K-1 份為訓練資料 1 份為測試資料，以利後續分析使用。

(二) 訓練階段：以交叉驗證方法，將資料分為 K-1 份訓練資料及 1 份測試資料。將 K-1 份訓練資料（轉換成矩陣圖的跟車車距資料）輸入 CNN 進行模式之訓練，經 CNN 分類駕駛員之跟車行為類型。因使用交叉驗證方法，必須重覆進行此步驟 K 次。

(三) 測試階段：以交叉驗證方法，將資料分為 K-1 份訓練資料及 1 份測試資料，將 1 份測試資料（轉換成矩陣圖的跟車車距資料）輸入已完成訓練的 CNN 模式，經 CNN 分類駕駛員之跟車行為類型。因使用交叉驗證方法，必須重覆進行此步驟 K 次，讓每一份資料都進行過驗證。

(四) 評估預測結果：利用精確度作為模型優劣評估指標進行模型準確度評估。

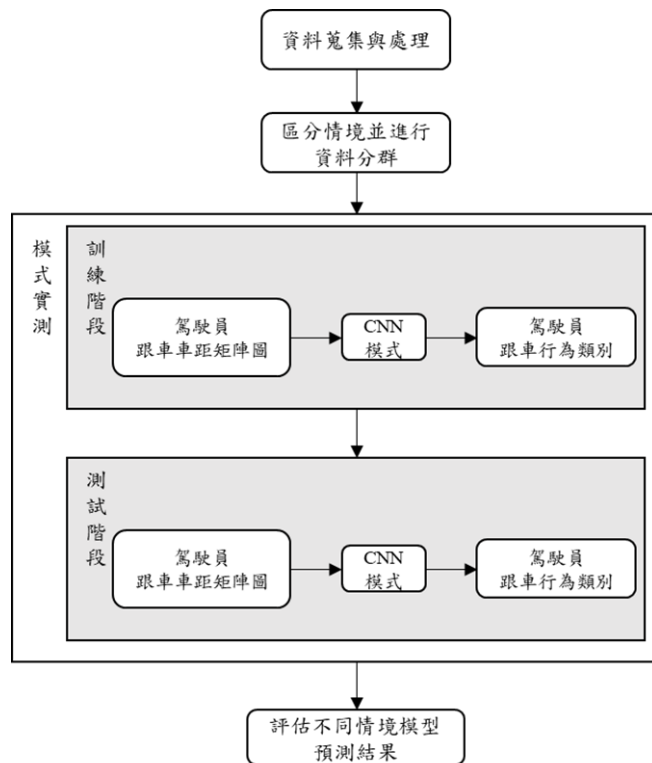


圖 4 研究架構圖

### 三、資料處理

由於須將 Mobileye 跟車車距資料化為矩陣圖，因此必須先將 Mobileye 之原始資料進行處理，Mobileye 原始資料之總資料筆數為 2,818,455 筆，本研究使用 R 語言對資料進行處理，處理流程如圖 5。

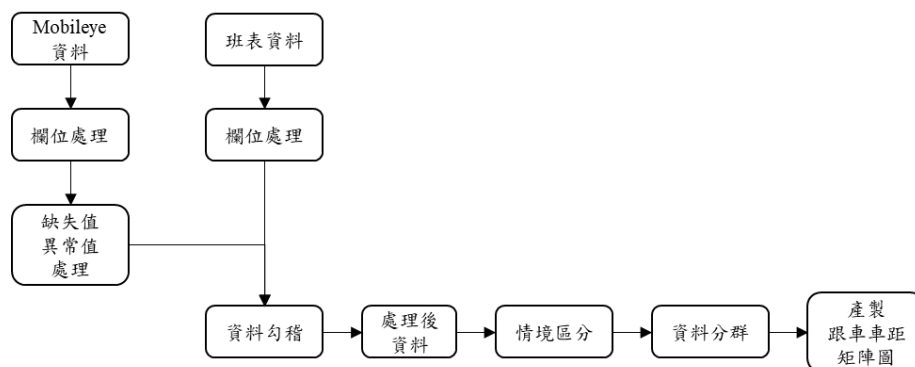


圖 5 資料處理流程圖

#### (一) Mobileye 資料

為將跟車車距資料轉換為矩陣圖，因此先將 Mobileye 原始資料之資料欄位進行篩選，僅留下相關欄位。而檢視 Mobileye 資料中，發現車速欄位中存在異常值，異常值之定義為不合理之車速，如 255km/hr，因此於本研究中視為異常值刪除該筆資料。此外，Mobileye 資料中之駕駛員編號存在缺失值，而本研究之目

的在探討駕駛員之跟車車距類型，因此先將駕駛員編號存在缺失值之資料予以保留，在後續資料處理步驟中與班表資料進行勾稽。經初步處理及彙整後，車速為異常值之筆數為 78 筆，駕駛員編號存在缺失值之筆數為 2,585,002 筆，異常值之資料予以刪除，而駕駛員編號缺失值資料則在資料勾稽步驟中進行處理。

## (二) 班表資料

為將跟車車距資料轉換為矩陣圖，將班表原始資料之欄位進行篩選。

## (三) 資料勾稽

由於 Mobileye 資料中之駕駛員編號存在缺失值，導致無法分辨該筆資料由何位駕駛員產生，故須將 Mobileye 資料與班表資料進行勾稽，將每筆資料對應至相符之駕駛員編號，資料勾稽步驟為，先對應車牌及資料日期，接著再確認資料時間落在班表區間為何位駕駛員駕駛之班次，並篩選出符合之駕駛員編號。

## (四) 處理後資料

經前述處理後之資料，如下表 2 所示。

表 2 處理後資料

| Sno     | CarNo  | SecUnit | Speed | UpdDayTime      | Employee_No |
|---------|--------|---------|-------|-----------------|-------------|
| 9017459 | 312-U8 | 101     | 38    | 2019/3/13 05:15 | B0250       |
| 9017458 | 312-U8 | 101     | 37    | 2019/3/13 05:15 | B0250       |
| 9017466 | 312-U8 | 101     | 27    | 2019/3/13 05:15 | B0250       |

## (五) 情境區分

由於本研究探討不同因素對於駕駛員跟車車距之影響，選定三個因素作為研究情境，分別為日夜、天氣及車速，並針對每筆資料進行情境分類。

1. 日夜因子：本研究為探討日夜因子是否影響駕駛員之跟車車距，因此參考中央氣象局「中華民國 108 年日出日沒時刻表」，依照日出及日落時間來區分情境，如：4 月 1 號日出時間為 05：46 日落時間為 18：10，若資料時間落在 05：46～18：10 之間即為日間，資料時間落在 05：46～18：10 之外即為夜間。
2. 天氣因子：本研究為探討天氣因子是否影響駕駛員之跟車車距，因此參考中央氣象局各測站之歷史資料，並選定本研究資料行駛之路線國道一號周邊之氣象測站，台北、五股、林口、中壢、楊梅、湖口、新埔及新竹八個測站，以發車時間之該小時時窗對應沿途行經氣象測站歷史資料之降雨與否，檢視該時窗之降雨資料，若八個測站中超過五個測站有降雨紀錄，則視為雨天，若無則視為非雨天，以此作為雨天與非雨天之判斷依據。
3. 車速因子：本研究為探討車速因子是否影響駕駛員之跟車車距，參考國道最外側車道最低限速 60 公里，而設定車速 60 公里為分界，Mobileye 資料中速度低於 60 公里即屬於車速慢；速度高於 60 公里即屬於車速快。

## (六) 資料分群

本研究欲產製之矩陣圖為  $28 \times 28$  大小之矩陣，此矩陣圖設置參考至 MNIST 手寫數字識別模型中輸入資料為  $28 \times 28$  大小之數字圖片。將前述步驟處理後之資料，針對每個駕駛員的所有跟車資料，將符合情境設定的資料篩出，並依照本研究設定矩陣大小  $28 \times 28$  根據篩出資料之時間序列連續取 784 個跟車車距數值資料放入矩陣中，以夜間情境資料為例，如圖 3-10 所示，最後剩餘之資料若不足 784 個則捨去，因每個駕駛員在各情境之資料量有所不同，因此一位駕駛員在每個情境中可能會有多個矩陣，如：駕駛員 A 在情境一中有 15349 筆資料，能產製 19 個矩陣，最終製成 19 張矩陣圖；在情境二有 5569 筆資料，能產製 7 個矩陣，最終製成 7 張矩陣圖。接著再將前述產製後的資料進行分群，將跟車資料以 K-means 來做分群並考慮不同情境及不同分類數量，使得每個情境中的所有資料都有對應的標籤，以利後續 CNN 模式分析。

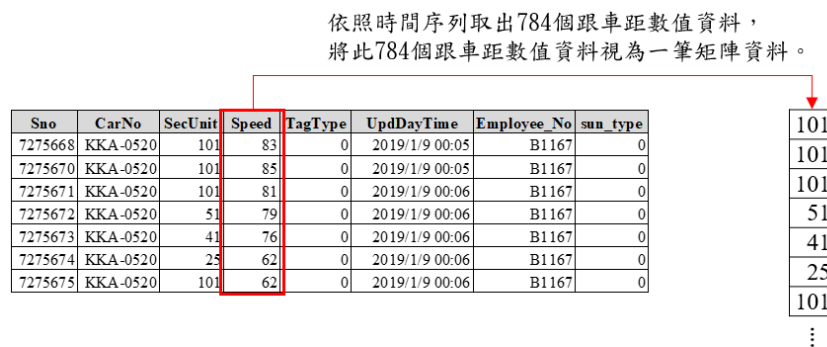


圖 6 跟車車距資料彙整示意圖

再將前述產出的所有矩陣資料以 K-means 進行分群，將檢視分群後資料來給定矩陣所屬之類型，如：積極型駕駛或保守型駕駛。

## (七) 產製跟車車距矩陣圖

本研究以 CNN 進行分析，而其輸入資料型態為影像，故在本研究中，將跟車車距矩陣資料視為一張張的影像資料，即為跟車車距矩陣圖。影像是由像素所組成，因此本研究將跟車車距矩陣的數值結合像素 (Pixel) 的概念轉換為圖片，圖 7 為本研究中之跟車車距矩陣圖，以顏色代表跟車距之遠近。

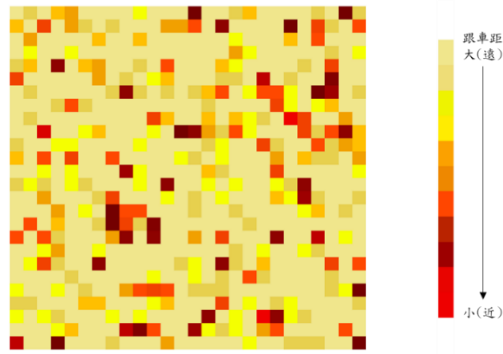


圖 7 跟車車距矩陣圖（一）

本研究在產製矩陣圖步驟中嘗試透過圖片預處理（Image pre-processing）的動作以避免過度擬合之情況，而圖片預處理是在模型訓練前事先對圖片進行處理，用來增加資料量或者避免模型訓練可能產生問題之方法。而其中資料增強（Data augmentation）為訓練 CNN 常用的手段之一，主要原理是透過改變圖片特徵位置並保留特徵與特徵間的關係，而常見的做法有：資料正規化、資料白化處理、影像處理等。本研究引入上述概念，欲強化各分群資料之特徵，在圖片預處理過程中，在不更改跟車車距數值資料的原則下，將每張圖片的跟車車距資料依照大小進行排列，圖 8 為經過圖片預處理後之跟車車距圖片，並嘗試以經圖片預處理的跟車車距矩陣圖進行實驗，測試是否影響辨識率。



圖 8 跟車車距矩陣圖（二）

#### 四、卷積神經網路架構

卷積神經網路（Convolutional Neural Network, CNN）是近年備受重視的人工智慧之一，廣泛被應用在影像辨識、人臉辨識、手寫辨識等多項應用，並獲得良好成效。其架構主要由三種不同類型的層連接所構成：卷積層（Convolutional Layer）、池化層（Pooling Layer）及全連接層（Fully Connection Layer）。

卷積層的主要目的，從輸入圖片中萃取出資訊，進一步產出特徵圖（Feature Map）。在卷積層中，透過不同的卷積核（Filter）來學習不同的特徵。卷積核為

$N \times N$  矩陣，其中的數字代表著不同的權重，依序對圖片中的矩陣由左上角掃描至右下角，並依序進行內積後，產出特徵圖。

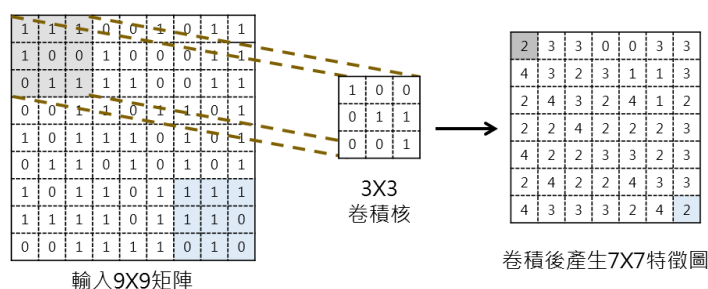


圖 9 卷積層示意圖

池化層通常會接在卷積層之後，池化的目的是將資料量減少並且保留其中的重要資訊，可以藉由取出池化範圍中的最大值、平均值或是加總來進行輸出資料尺寸縮小的動作。

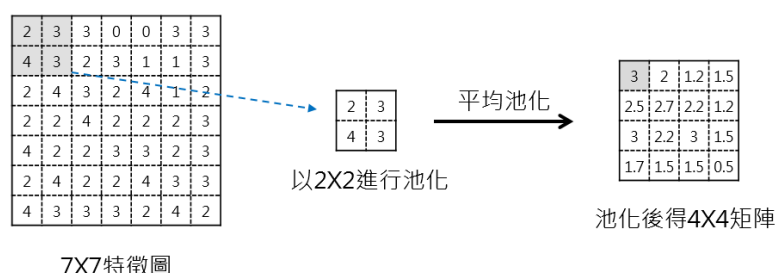


圖 10 池化層示意圖

全連接層為一般的類神經網路，主要目的是將經過多層卷積及池化後的特徵資料進行分類，透過激活函數得到判斷出的機率值。全連接層必須將前一步驟的產出結果進行平坦化轉換為一維資料，在全連接層中，每個神經元都與上一層之神經元彼此相連接，通常會設定兩層的全連接層，而各個連結都有獨立且相異的權重值，對各種分類情況輸出的機率值進行分類，最後輸出預測結果。

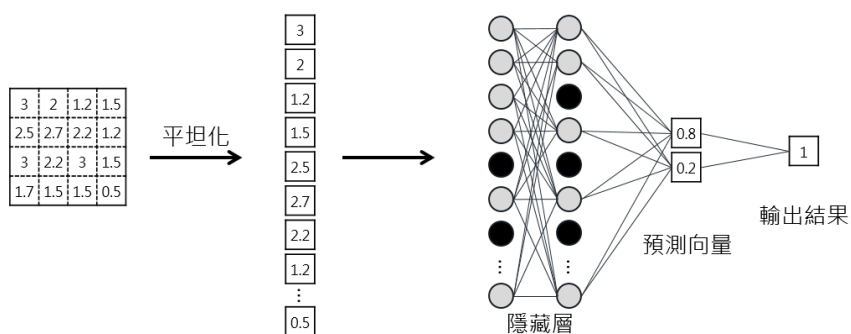


圖 11 全連接層示意圖

## 肆、實驗結果與分析

本研究利用深度學習套件 MXNet 進行 CNN 模式實測，並利用 R 語言 (Version 3.5.1) 編寫程式碼。2018 年版 ASUS VivoBook S14 (Intel Core i5-8250U CPU、Intel UHD Graphics 620 GPU、8GB RAM) 為本研究使用之硬體環境。

並以 HT 客運公司裝設於車輛上之車載系統 Mobileye 所蒐集之跟車車距資料，以各個駕駛員之實際跟車車距資料為基礎並使用 CNN 模型分析駕駛員之跟車行為類型，而本研究期望能發展出一套可分類駕駛跟車行為類型且可靠的分類模式，故本研究將提出不同情境組合進行測試比較，並分析各情境的測試結果。

### 一、模式與參數設定

本研究所使用之卷積神經網路模型參考 Ma et al (2017) 利用 CNN 模式預設號誌化道路速率之模型架構、Yann (1998) 提出之 CNN 模型及 MNIST 手寫數字識別模型，根據前述 CNN 作為模型設置參考依據。本研究之 CNN 設置為兩層卷積層、兩層池化層及兩層全連接層所構成之模型架構，本研究 CNN 之詳細設置及參數如表 3 及表 4。

表 3 卷積神經網路設置及參數

| 模式架構    | 說明                                                                | 維度           |
|---------|-------------------------------------------------------------------|--------------|
| 輸入資料    | 跟車車距矩陣圖 $28 \times 28$                                            | (1, 28, 28)  |
| 第一層卷積層  | 卷積核： $2 \times 2$<br>20 個卷積核                                      | (20, 25, 25) |
| 第一層池化層  | 使用平均池化法 (mean-pooling)<br>卷積核： $2 \times 2$<br>移動步長： $2 \times 2$ | (20, 13, 13) |
| 第二層卷積層  | 卷積核： $2 \times 2$                                                 | (50, 12, 12) |
| 第二層池化層  | 使用平均池化法 (mean-pooling)<br>卷積核： $2 \times 2$<br>移動步長： $2 \times 2$ | (50, 6, 6)   |
| 第一層全連接層 | 500 個隱藏層                                                          | (500,)       |
| 第二層全連接層 | 2 個隱藏層<br>SoftmaxOutput 函數達到分類目的                                  | (2,)         |
| 輸出      | 分類結果                                                              | (2,)         |

表 4 卷積神經網路相關參數

| 參數   | 說明                                                                                    |
|------|---------------------------------------------------------------------------------------|
| 退出比率 | 卷積神經網路的全連接層是由數個神經元組成，而為了避免訓練過程中過度擬合的情況，因此設置退出層來避免過度擬合，而退出比率通常設置為 0.5，意即隨機放棄 50% 的神經元。 |
| 訓練次數 | 將訓練集中所有的數據進行一次完整的訓練。                                                                  |
| 輸入批量 | 在訓練資料樣本數較龐大的情況下，必須將訓練資料分批輸入模型中訓練，本研究輸入批量為 300，此設定係經由本研究測試得到較佳之設定值。                    |
| 學習率  | 學習率影響權重更新之速度，通常建議設定為 0.01~0.001 間，本研究設定為 0.001。                                       |

本研究為防止模型在實測時有過度擬合之狀況，因此在模型中使用 K 折交叉驗證 (K-fold cross-validation)，將資料分為 K 份，每次選擇 K-1 個為訓練集其餘 1 個為測試集，重複 K 次並將 K 次的平均交叉驗證正確率作為實證結果。本研究將資料集以 3：1 的比例分為訓練集及測試集，再做四次交叉驗證 (4-fold cross-validation)，如圖 12 所示。

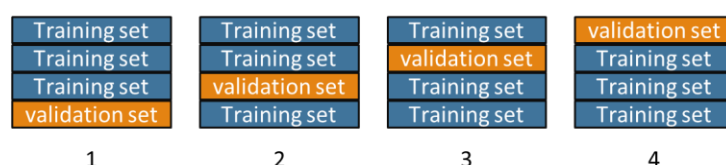


圖 12 四次交叉驗證 (4-fold cross-validation) 示意圖

## 二、情境設定

本研究考量三大因子進行實測，分別為天氣因子、車速因子及日夜因子，以下針對三大因子及各種情境組合說明，如

表及

表所示。

表 5 實測因子條件表

| 因子 | 條件說明          |
|----|---------------|
| 天氣 | 非雨天           |
|    | 雨天            |
| 車速 | 時速大於 60 km/hr |
|    | 時速小於 60 km/hr |
| 日夜 | 日間            |

|  |    |
|--|----|
|  | 夜間 |
|--|----|

根據前述因子條件，本研究建構以下七種實測情境，以探討在不同實測情境組合下之分類結果差異，各情境組合詳細內容如

表。

表 6 實測情境表

| 實測情境編號 | 考量因子 | 分析因子                     | 情境說明                                                                                             |
|--------|------|--------------------------|--------------------------------------------------------------------------------------------------|
| 情境一    | 單一因子 | 日夜                       | 僅考慮日、夜條件下對駕駛員跟車車距之影響。                                                                            |
| 情境二    |      | 天氣                       | 僅考慮雨天、非雨天條件下對駕駛員跟車車距之影響。                                                                         |
| 情境三    |      | 車速                       | 僅考慮車速快、慢條件下對駕駛員跟車車距之影響。                                                                          |
| 情境四    | 兩因子  | 日夜<br>×<br>車速            | 考慮日夜及車速因子產生之四種情境組合對駕駛員跟車車距之影響：<br>1.日間及車速快<br>2.夜間及車速快<br>3.日間及車速慢<br>4.夜間及車速慢                   |
| 情境五    |      | 天氣<br>×<br>車速            | 考慮車速及天氣因子產生之四種情境組合對駕駛員跟車車距之影響：<br>1.非雨天及車速快<br>2.雨天及車速快<br>3.非雨天及車速慢<br>4.雨天及車速慢                 |
| 情境六    |      | 日夜<br>×<br>天氣            | 考慮日夜及天氣因子產生之四種情境組合對駕駛員跟車車距之影響：<br>1.日間及非雨天<br>2.日間及雨天<br>3.夜間及非雨天<br>4.夜間及雨天                     |
| 情境七    | 三因子  | 日夜<br>×<br>天氣<br>×<br>車速 | 考慮日夜、車速及天氣三因子產生之八種情境組合對駕駛員跟車車距之影響：<br>1.日間、非雨天及車速快<br>2.日間、雨天及車速快<br>3.夜間、非雨天及車速快<br>4.夜間、雨天及車速快 |

|  |  |  |                                                            |
|--|--|--|------------------------------------------------------------|
|  |  |  | 5.日間、非雨天及車速慢<br>6.日間、雨天及車速慢<br>7.夜間、非雨天及車速慢<br>8.夜間、雨天及車速慢 |
|--|--|--|------------------------------------------------------------|

### 三、各分群之各情境實驗結果

本研究為將客運駕駛員之跟車行為類型分類，以 K-means 集群分析方法將駕駛員跟車車距資料分為兩群以及三群，將資料分為積極型駕駛及保守型駕駛，分群完畢後再將跟車車距數值矩陣轉換為跟車車距矩陣圖輸入 CNN 模型中。

以本研究所設定之情境一為例，考慮日夜因子作為跟車車距矩陣圖產製依據之情境，根據每筆跟車車距資料之時間分類為日間資料及夜間資料，再依據資料之時間序列依序匯入矩陣中，進一步產製跟車車距矩陣圖，並將情境一之所有跟車車距矩陣圖一併輸入 CNN 中分析，步驟流程如圖(1)。為探討是否能提升分兩群情形下之模型分類能力，本研究參考資料增強概念，在輸入模型前先對矩陣圖進行圖片預處理動作，將每個數值視為一個特徵並在不更改跟車車距數值資料的原則下，透過改變圖片特徵位置的手段來達到資料增強之目的，並針對本研究設定之七個情境進行實驗，步驟流程如圖(2)。

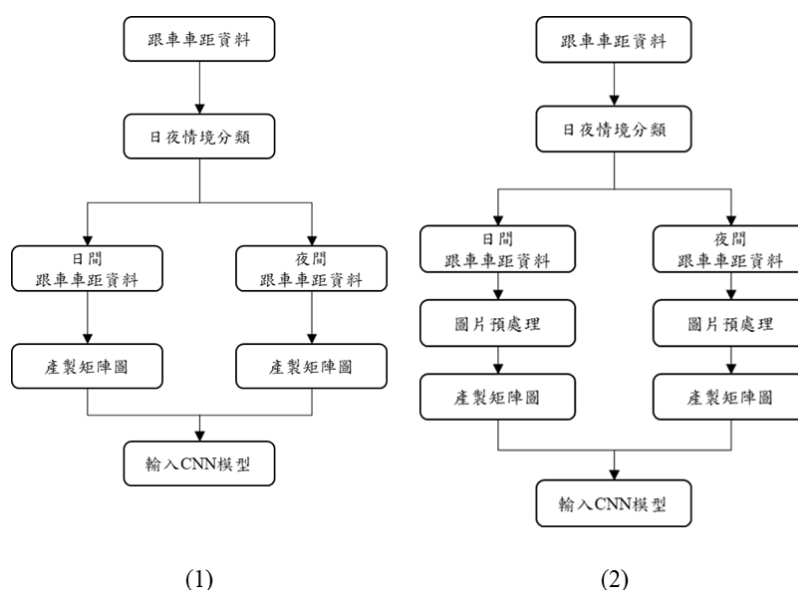


圖 13 實驗步驟流程圖

各情境實驗結果，係由四次交叉驗證之測試精確率而得，其中以 K-means 分兩群在七個情境下的實驗結果如圖 14。探討經 K-means 分群將跟車車距矩陣圖分為兩類的情形下之實驗結果，並探討每一個情境的模型分類能力，進一步嘗試透過圖片預處理步驟檢視實驗結果的影響，由圖 14 證實經圖片預處理之資料模

型分類能力平均為 97%，相較於未經圖片預處理之結果，其分類能力明顯提升，且具有十分可信的辨識能力。

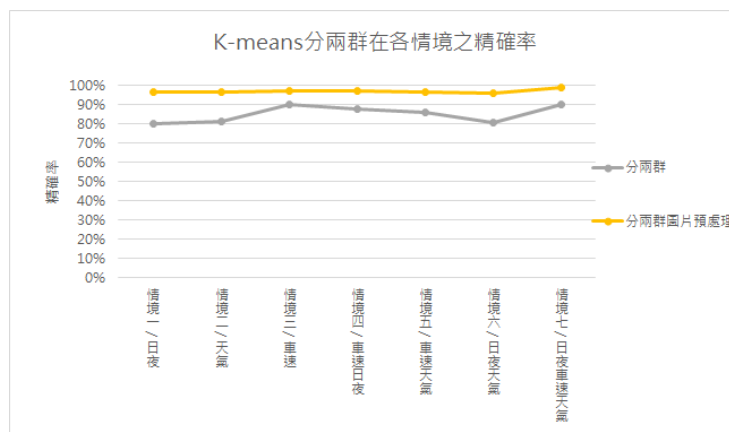


圖 14 K-means 分兩群實驗結果

接著嘗試以 K-means 將跟車車距矩陣圖分為三類，並探討每一個情境的模型分類能力，進一步嘗試透過圖片預處理步驟以提升模型分類能力，由圖 15 可知僅分三群之資料模型分類能力平均 38%，經圖片預處理之資料模型分類精確度皆有超過五成之分類能力，經圖片預處理步驟確實讓每個情境分類能力提升。

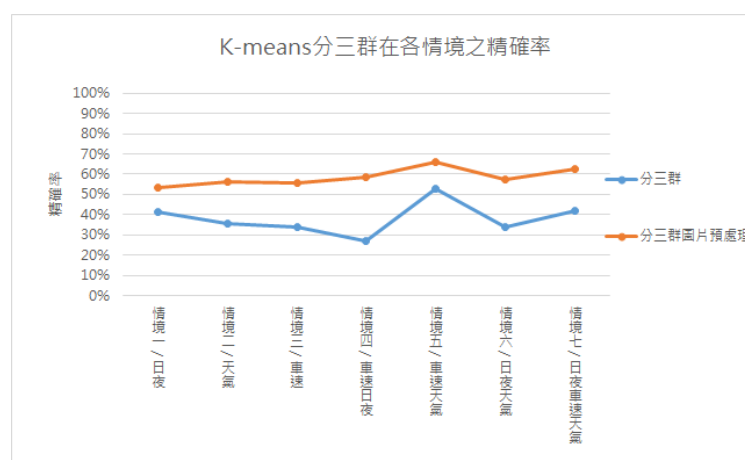


圖 15 K-means 分三群實驗結果

## 伍、結論與建議

本研究主要以 CNN 來建構一套駕駛跟車行為類型分類模型，以客運業者提供之 Mobileye 跟車車距資料作為輸入資料，並設計日夜、天氣及車速三個因素來篩選資料，以 K-means 將輸入資料分別分為兩類及三類，探討不同分類情形之模型分類能力。利用本研究產製之跟車車距矩陣圖進行實驗後，發現本研究所提出之模型在各個情境下都有不錯的預測成效，以下將根據本研究之成果做結論並提出未來相關研究之建議。

## 一、結論

(一) 本研究利用 HT 客運公司之 Mobileye 車載系統所記錄之跟車車距資料並搭配天氣、車速及日夜三因素進行篩選，依不同情境設定產製出之跟車車距矩陣圖作為輸入 CNN 之輸入資料，並輸出該圖片的分類預測。

(二) 本研究經由實驗分析後發現以 K-means 將跟車車距矩陣圖分為積極型及保守型兩類，相較於分為積極型、穩健型及保守型三類的情形，分兩類有較佳的預測結果，與李佩憶（2017）提出以集群分析將偏差駕駛行為分為兩群為較佳的分群數量相同，此分類數量可作為後續應用之參考。

(三) 本研究證實經圖片預處理步驟確實提升分類精確率，此類型之資料在輸入模型前可參考圖片預處理方法。

(四) 本研究探討天氣、車速及日夜三因素作為篩選依據，並進一步分別探討單一因子、同時考慮兩因子及考慮三因子之情境對模型精確率之影響，在實驗結果中發現隨著考慮因素的增加，模型精確率也會隨之上升，如圖 16 所示，除了 K-means 分三群同時考慮兩因子的情境五（車速天氣）以及 K-means 分兩群僅考慮單一因子的情境三（車速），這兩類精確率高於考慮三因子的情境七，其餘情境皆隨著考量因素的增加，而提升精確率。

(五) 經由本研究實驗結果，模型分類精確率以 K-means 分兩群且進行圖片預處理的情況為最佳，七個情境的分類精確率平均高達 97%。其餘的結果中，精確率最高為 K-means 分兩群僅考慮車速因子的情境三，其精確率高達 90%；以及 K-means 分兩群考慮三因子的情境七，其精確率高達 90%；並發現在考慮車速及天氣的情境下，都具有較高之模型分類能力，車速與天氣兩因子可能為影響駕駛員跟車車距行為的重要因素。

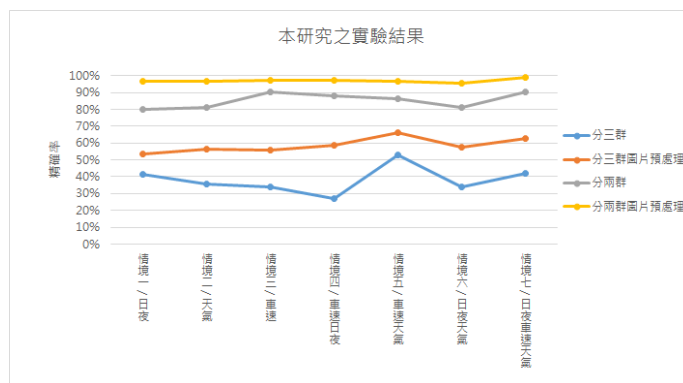


圖 16 所有情境組合實驗結果

## 二、建議

(一) 本研究使用之跟車車距資料所記錄之時間間隔並非等距，意即每筆資料時間間距不一致，若未來有機會取得回傳頻率相同之資料，即可以一段時間為單位進行分析，觀察每段時間駕駛員的跟車行為變化，如：取得每兩秒回傳一筆

之原始資料，即可將輸入矩陣設置為以 15 分鐘為單位產製之矩陣圖，意即此張矩陣圖中含有的資料為 15 分鐘內的連續跟車資料。

(二) 本研究輸入資料處理過程中，判定天氣條件之依據選用離高速公路較接近之八個氣象測站每日降雨歷史資料來推估是否降雨，並以客運發車時間為時窗作為推斷之依據，因此可能存在不完全準確之問題。建議未來可結合定位資料來推斷資料之該時刻降雨與否，進行綜合判斷作為推斷雨天及非雨天之依據。

(三) 本研究所使用之跟車車距資料並未考慮當下之交通情況，交通情況可能為影響跟車情形的因素之一，如：平日及周末、是否有交通事故、道路施工及特殊節慶等，皆可能對分析結果造成些微影響，因此建議未來可將相關影響因素考慮至預測模式內，以建立更完善的模式。

(四) 本研究在資料處理步驟僅考慮車速及跟車車距資料，並未採用其他資料，建議未來進行相關研究，若能取得 Mobileye 智能攝影機之影像，也可更清楚每個跟車資料當下情況。

(五) 本研究未考慮與駕駛員身、心理之相關因素，Lianzhen & Yulong. (2014) 對公路客運駕駛員主觀疲勞程度進行量測，發現不同長度的駕駛時間需要有不同的休息時間，才能使駕駛員注意力及反應力回復。建議未來可藉由訪談得知駕駛員最易產生倦意之時段並結合班表所排之開車時段與休息時段進行探討，進一步分析該時段的跟車行為是否有異常，若有是否可透過休息時間來改善。

## 參考文獻

- 交通部統計處 (民 105)，「104 年民眾日常使用運具狀況調查—摘要分析」
- 吳明隆 (民 107)，「SPSS 操作與應用—多變量分析實務」，台北。五南書局。
- 李佩憶 (民 106)，「人格特質對於駕駛行為特性影響與駕駛風險分級制度之研究—以國道客運駕駛員為例」，國立成功大學交通管理科學系碩士班碩士論文。
- 黃冠維 (民 107)，「利用卷積神經網路預測市區道路路況之研究」，國立交通大學運輸與物流管理學系碩士論文。
- 陳信丞 (民 103)，「國道客運公司駕駛員人因特性、車輛駕駛行為與駕駛績效之研究」，國立成功大學交通管理科學系碩士班碩士論文。
- 警政統計年報 (民 103)，「道路交通事故 (A1) 類原因、傷亡」，內政部警政署。
- Feng, S., Li, Z., Ci, Y., and Zhang, G. (2016). Risk factors affecting fatal bus accident severity: Their impact on different types of bus drivers. *Accident Analysis and Prevention*, Vol. 86, pp. 29-39.
- Milan Vujanić, Boris Antić, Dalibor Pešić, Milan Savićević, (2016), Testing the psychophysical characteristics of professional drivers – Can we identify unsafe drivers?, *Transportation Research Part F: Traffic Psychology and Behavior*. Volume 42, Part 1, pp 104-116.
- Ma, X., Dai, Z., He, Z., Ma, J., & Wang, Y. (2017). Learning traffic as images: A Deep Convolutional Neural Network for Large-Scale Transportation Network Speed Prediction. *Sensors*, 17 (4), 818.
- Pack, A.M., Rodgman, E., Cucchiara, A., Dinges, D.F., Schwab, C.W. (2005), Characteristics of crashes attributed to the driver having fallen asleep, *Accident Analysis & Prevention*, Vol 27, pp. 769-775.
- Lianzhen, W., Yulong, P. (2014), The impact of continuous driving time and rest time on commercial drivers' driving performance and recovery, *Journal of Safety Research*. Vol 50, pp. 11-15.
- U.S General Services Administration Fleet Drive-thru, Avoid the Six Most Unsafe Driving Behaviors.