

高速公路道路幾何及環境與交通事故之影響研析 ——以中區為例

呂文玉¹、張洺瑋²、劉欣憲³、吳國銓⁴

摘要

為增進道路安全，各國皆透過系統性方法分析公路基礎設施的安全風險，其中，國際道路評估計畫（iRAP）便是採用科學化、證據導向的評估方法，協助各國找出高風險道路、提出改善方案。iRAP 評估方法是有 5 個步驟，包括事故風險地圖、星級評分、死亡重傷估算、道路投資計畫及績效追蹤的方式，透過模型與實地調查進行評估及改善。以星級評分為例，其針對每 100 公尺路段，對各道路屬性逐項編碼，以標準化分類，對不同用路人給予不同的風險權重，再將風險權重加總組成星級評分分數（Star Rating Score, SRS），再依分數將道路對應 1 至 5 星。

本文參考 iRAP 之精神，針對高公局中分局轄區路段切分為每公里，並包含道路幾何線型、環境變數、交通管制措施、車流量 4 大類之道路屬性，共 24 種道路變數。結合 110-113 年公警全般事故資料，透過逐步迴歸、XGboost（eXtreme Gradient Boosting）、隨機森林（Random Forest）等機器學習方法進行分析，希冀藉由分析成果找出事故影響之變數組合，以提出具體有效之建議改善策略，進一步降低交通事故數量。

關鍵詞：道路環境變數、逐步迴歸、XGboost。

一、前言

高速公路作為現代交通運輸的關鍵基礎設施，其運行安全直接關係到社會經濟發展與公眾人身安全。由於車輛行駛速度快、交通流量大，高速公路上的交通事故往往具有較高的嚴重性，導致人員傷亡與財產損失。然而，道路的安全性受到多種因素的影響，其中環境條件在事故發生與嚴重程度上扮演關鍵角色。道路的幾何設計、操作條件以及天氣變化，皆可能改變駕駛行為並影響車輛的穩定性。

為研析道路環境變數和事故之影響，本文參考各國案例，其中最早為歐洲道路評估計畫（European Road Assessment Programme, EuroRAP）透過系統性的方法對道路幾何進行分類與分級，透過道路幾何（如曲線半徑、坡度）、車道和路肩寬度、路側設施、路面狀況等道路環境，根據交通量及事故件數，綜合判斷

¹ 交通部高速公路局中區養護工程分局分局長。

² 交通部高速公路局中區養護工程分局幫工程司，臺中市西屯區臺灣大道四段 5 巷 55 號，04-22529181 分機 2103，C00373@freeway.gov.tw。

³ 揚立科技有限公司總經理。

⁴ 國立屏東大學資訊工程學系副教授。

道路安全等級，以利歐洲各國政府作為道路加強作為及成本效益評估的依據。這套模式進而發展成國際道路評估計畫（International Road Assessment Programme, iRAP），也就是 EuroRAP 的全球延伸，並已於 132 個國家、190 萬公里的道路和設計進行評估，並繪製道路風險地圖。

本文參考 iRAP 之道路評估精神，針對高公局中分局轄區路段，包含國道 1 號 100.8-251.1k、國道 3 號 110.7-270k、國道 4 號全線、國道 6 號全線，總長度約 373.967 公里之路段資料，切分為每公里，並包含道路幾何線型（R 值、A 值、橫向坡度等）、環境變數（逆光、照明、橋梁、隔音設施等）、交通管制措施（常態開放路肩、牌面、測速照相、速限等）、車流量（路段年車流量）4 大類之道路屬性，共 24 種道路變數資料。為減少無用的道路環境變數或其變數間互相產生之共變效應，進而影響模型的解釋效果，另也為減少車流量對於事件數所產生之過度配適（overfitting），依據本文分析重點，將 4 大類的變數依序透過逐步迴歸分析之投入，分析篩選顯著的道路環境變數。將顯著的道路環境變數分別以極限梯度提升（eXtreme Gradient Boosting, XGB）、隨機森林（Random Forest, RF）之機器學習的方式，了解其特徵重要性，並進一步查看特徵重要性排名順序是否一致，以及特徵性的正負相關性是否與逐步迴歸分析之係數相符。再行將分析結果之重要道路幾何條件或環境變數，透過交通工程加強作為之實作方式，增進其道路安全。

二、文獻回顧

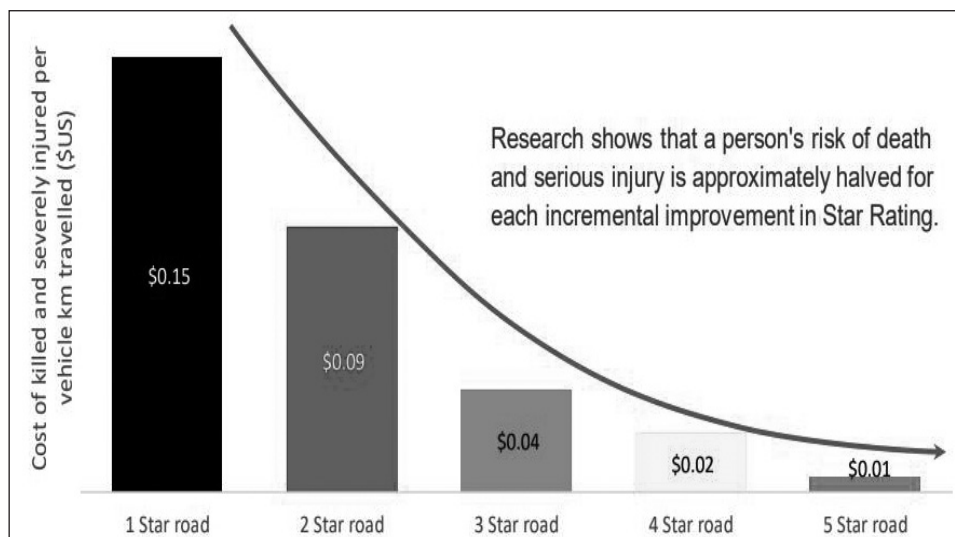
2.1 道路環境與交通事故的影響分析

許多研究顯示，道路環境與事故率之間存在顯著關聯，然而，這些因素的影響力大小不一，且可能與駕駛行為交互作用。傳統經常利用統計方法來分析交通事故，隨著 AI 和大數據的盛行，越來越多的研究透過這些新的技術進行分析，為深入了解道路環境對事故的影響，分析道路事故與道路環境變數之間的關係。以下分成「風險評估」、「統計模型」及「機器學習」三部分的相關文獻進行文獻探討。

2.1.1 風險評估

國際道路評估計畫（International Road Assessment Programme, iRAP）係透過每 100 公尺路段中 52 種道路屬性，包括有無人行道、路側設施、車道數量、速限、車流量等，並針對每種事故類型（如追撞、對撞、行人衝突等）都有一個事故修改因子（Crash Modification Factors, CMF）。透過每個屬性對該事故之事故修改因子進行加總，計算出星級分數（Star Rating Score, SRS），依據分數分配 1~5 星。

多項研究已證明，星級評分與死亡及重傷事故成本之間存在直接關係。每當星級評分提升一級，死亡和重傷所造成的成本通常會減少一半，如圖 1。



參考資料：iRAP Star Rating and Investment Plan Manual

圖 1 星級評分與死亡及重傷事故成本之間的關係

2.1.2 統計模型

為了解各國如何應用傳統的統計模型，分析道路幾何與環境因素變數如何影響高速公路的事故件數及頻率，本文彙整各國研究文獻如表 1。相關研究顯示道路的環境狀態會影響事故發生頻率及嚴重程度，惟部分道路環境變數影響程度較小。

表 1 應用傳統統計模型文獻彙整表

作者	研究目標	研究方法	分析資料	結論
Venkataraman Shankar 等人 (1995)	道路幾何與環境因素如何影響鄉村高速公路的事故頻率	負二項迴歸模型	1. 美國 I-90 高速公路 1988 年至 1993 年事故數據 2. 道路幾何變數（如彎道數量、彎道半徑、坡度） 3. 環境變數（如降雨量、降雪量和氣候變化）	1. 當道路曲率較高時，事故頻率顯著增加。 2. 坡度超過 2% 的區域也顯示出較高的事故發生率，尤其是在降雪條件下，坡度與事故風險呈正相關。 3. 單日降雨量超過 2.54 cm 或降雪量超過 5.1 cm 時，事故發生率顯著上升，並且降雨與彎道、降雪與坡度的交互作用使得事故風險進一步提升。
Kopelias, P. 等人 (2007)	道路幾何特徵、操作條件及天氣因素對事故發生與嚴重程度的影響	皮爾森相關分析、變異數分析、線性迴歸模型	1. 2004 至 2005 年 Attica Tollway 之 533 起碰撞事件數據，分為總體數據和個別事故數據兩類。 2. 道路幾何：車道數量、車道寬度、坡度、曲率半徑。 3. 操作因素：速度限制、匝道距離、交通流量。 4. 天氣條件：降雨、路面濕滑程度。 5. 事故嚴重程度：根據傷亡與財物損失分類。	1. 道路幾何設計、速度限制與天氣條件對碰撞率有影響，但駕駛行為仍是主要影響因素。 2. 低速限制區域、匯入 / 匯出區域、坡度較大的道路與急彎處事故率較高。 3. 雨天與濕滑路面雖會增加事故風險，但整體影響較小，僅能解釋約 5% 至 10% 的事故發生。 4. 應將重點放在駕駛行為管理與違規監測，以提升高速公路的安全性。

作者	研究目標	研究方法	分析資料	結論
Chun Yip Chan 等人 (2009)	道路幾何形狀與環境因素如何影響鄉村高速公路的事故頻率	負二項迴歸模型、交叉表分析	2006 年田納西州 110 英里的城市高速公路事故數據。 - 國際平整度指數 (IRI)、當前服務性指數 (PSI)、車轍深度 (RD)、年平均日交通量 (AADT)、路肩寬度 (RS)、中央分隔帶距離 (LC)，以及不同類型的交通事故頻率。	- PSI 和 IRI 是影響事故頻率的關鍵變數，PSI 低於 3 會顯著增加事故風險，而較高的 IRI 也與事故頻率正相關。 - RD 在模型中未顯示顯著影響。 - 側撞事故與低 PSI 相關，而後車追撞事故則與較深的 RD 有關。 - 較寬的路肩與中央分隔帶距離可有效降低事故頻率。 - PSI 被認為是最佳的事務預測指標，可作為道路維護與交通安全改善的依據。
Xiaofei Wang 等人 (2023)	高速公路垂直曲線組合 (sag 曲線) 的幾何屬性對事故頻率的影響	泊松迴歸模型、負二項迴歸模型、階層泊松模型、階層負二項模型、貝葉斯推論	2011-2017 年華盛頓州六條高速公路 157 個 sag 曲線組合路段的數據 - 事故記錄、交通流量 (年平均日交通量 AADT)、道路幾何設計參數 (水平曲率、垂直曲率、坡度、曲率比值、錯位佈局等)，以及道路配置特徵 (車道數、路肩寬度、速限等)。	- 水平曲率、垂直曲率、出彎坡度、曲率比值及前置錯位等幾何特徵對事故頻率具有顯著影響。 - 較高的水平與垂直曲率、負出彎坡度、較大的曲率比值會增加事故風險，而前置錯位的設計可能導致駕駛員錯誤判斷。 - 較高限速 (65 或 70 英里 / 小時) 路段的事故率較低，可能與更高的道路設計標準有關。 - 研究結果用於優化高速公路幾何設計，提高道路安全性。

2.1.3 機器學習

為進一步如何應用機器學習分析此議題，並和傳統統計模型進行比較，本文彙整各國研究文獻如表 2。相關研究顯示機器學習模型在預測事故嚴重程度及事故事件優於傳統統計模型，其中 XGB 和 RF 表現較佳，另其能考慮道路環境變數交互作用之影響。

表 2 應用機器學習文獻彙整表

作者	研究目標	研究方法	分析資料	結論
Xiao Wen 等人 (2022)	道路幾何與環境因素如何影響高速公路的 ROR 事故	機器學習模型 (CART、MARS、RF、LightGBM) 進行碰撞頻率建模，並應用多種解釋性分析方法 (LIME、LSA、PDP、GSA、SHAP) 評估模型可解釋性。	- 美國華盛頓州 2014 至 2018 年間發生於州際公路和高速公路上的偏離道路 (Run-off-road, ROR) 事故數據 - 道路幾何變數 (彎道曲率、車道數、路肩寬度)。 - 交通流量 (每車道平均年日交通量)。 - 環境變數 (限速、坡度)。	- 限速、彎道曲率、車道數及交通流量等因素對 ROR 事故的發生具有顯著影響，且其影響呈現非線性關係。 - SHAP 分析顯示，高速限速與高曲率彎道的結合會顯著提高碰撞風險。 - ML 模型能有效預測事故風險並提供修正因子估算，但無法直接區分因果關係，未來可結合因果推論技術提升解釋性與應用價值。

作者	研究目標	研究方法	分析資料	結論
J.P.S. Shashiprabha Madushani 等人 (2023)	道路幾何與環境因素如何影響高速公路的事故風險	邏輯迴歸、決策樹、隨機森林、K-近鄰、XGBoost、SHAP、LIME	1. 斯里蘭卡高速公路 2017-2021 年事故數據，最終篩選 279 筆進行分析。 2. 環境變數包括道路表面狀況（乾燥 / 濕滑）、天氣條件（晴天 / 陰天 / 雨天）、照明條件（白天 / 夜間有良好或無照明）。 3. 道路幾何變數包括結構（直線 / 曲線）、事故發生地點（交流道 / 非交流道）。	1. 機器學習模型在預測事故嚴重程度方面優於傳統的邏輯迴歸，其中 XGB 和 RF 表現最佳。 2. SHAP 和 LIME 分析表明，天氣條件和路面狀況是影響事故嚴重程度的主要因素，惡劣天氣與濕滑路面顯著增加嚴重事故的風險。 3. 可解釋機器學習能有效提升模型透明度。
Yang Yang 等人 (2022)	道路環境變數影響高速公路的事故嚴重程度	XGBoost、SHAP、貝葉斯網路、剪枝樹傳播	1. 中國河北省 2018 年高速公路交通事故數據，共 567 筆事故資料。 2. 35 個屬性變數，最終篩選出 11 個影響事故嚴重程度的關鍵變數。 3. 道路類型、中央隔離設施、道路側邊防護設施、光照條件、事故發生時間、地形、天氣等。	1. 道路側邊防護設施、道路區段類型、中央隔離設施、光照條件與事故發生時間是影響事故嚴重程度的主要因素。 2. 考慮變數交互作用的 XGBoost-貝葉斯網路模型相較於僅考慮單一變數影響的模型，能夠更準確地預測事故嚴重程度，並提升預測效能。 3. 模型的預測準確率達 89.05%。
Jianyu Wang 等人 (2023)	道路環境變數及駕駛人狀態影響事故嚴重程度	隨機森林、SHAP、Apriori	1. 中國瀋陽市 2018-2020 年間的交通事故數據，共 2022 筆。 2. 19 個屬性變數。 3. 道路屬性（道路類型、物理隔離設施、道路網密度等） 4. 環境屬性（天氣、季節、商業與生活服務設施密度等） 5. 時間屬性（是否為高峰時段、是否為夜間等） 6. 肇事者特徵（年齡、駕駛經驗、責任類型等） 7. 事故特徵（碰撞類型、事故發生地區、事故原因等）	1. 交通事故嚴重程度的關鍵因素包括道路類型、季節、肇事者出行方式、事故發生區域、物理隔離設施、商業與生活服務設施密度等。 2. 郊區與低等級道路的事故更容易導致嚴重後果，而在人流密集的商業區雖事故發生率較高，但因車速較低，事故嚴重程度相對較低。 3. 冬季因低溫與降雪導致事故嚴重程度增加。 4. 本研究的機器學習與關聯規則分析方法提供了更具解釋性的交通風險因素識別。

2.1.4 小結

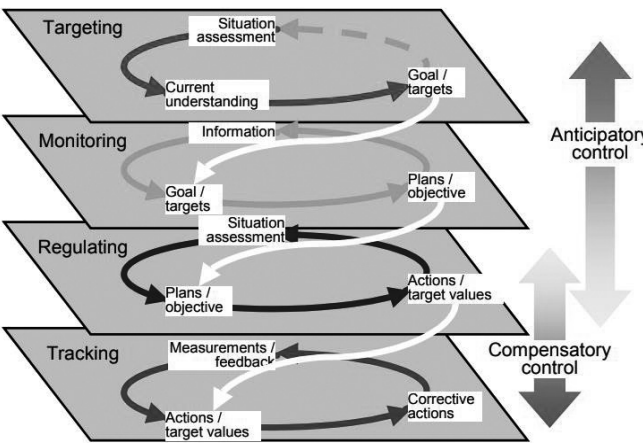
綜合上述文獻探討，各國皆致力應用「風險評估」、「統計模型」及「機器學習」等不同方法分析道路幾何、環境變數及相關事故特徵資料分析事故發生頻率和事故嚴重程度之關係。

本文參考 iRAP 風險評估之精神，將路段資料切分為每公里，並參考傳統統計模型分析此議題之限制，利用其方式作為篩選變數之依據，再行透過機器學習的方法，考慮道路環境變數交互作用，討論其對於事故發生頻率之影響。

2.2 事故改善作為探討

駕駛的決策過程是一個連續的動態過程，需要不斷感知環境訊息並進行判斷與選擇。Hollnagel (2003) 提出駕駛控制模型 (Driver-in-Control Model)，將駕駛表現由下到上分成四種模式，分別為追蹤 (Tracking)、調節 (Regulating)、監

控（Monitoring）和目標（Targeting），如圖 2 所示。除駕駛行駛決策外，可透過調整高速公路道路環境變數、道路標誌等，提升整體道路環境，改善路面幾何結構、道路安全性、號誌管理、智能交通系統及環境條件影響管理等層面，進而降低事故發生率或嚴重程度。



參考資料：Hollnagel et al., 2005

圖 2 駕駛控制模型

美國聯邦公路管理局（FHWA）透過大量研究交通安全研究機構、國際交通安全組織等各種專業報告、數據分析及全球經驗，於 2008 年起發布《有效安全對策手冊》（Proven Safety Countermeasures initiative，PSCi），考量路面設計結構、交通管理、道路安全性、智能交通系統及環境條件影響管理等層面，提出 28 項已驗證的有效安全對策，本文擷取對於高速公路交通安全有用之對策如表 3 所示。

表 3 擷取美國聯邦公路管理局《有效安全對策手冊》對策彙整表

類別	項目	說明
速度管理	測速相機	運用設置測速相機，彌補傳統執法方法、工程措施及教育
	可變速限	依據道路交通量、天氣及路面狀況等，顯示適當限速給駕駛者，以提高能見度、提供額外停車時間、減少衝擊力或降低二次事故發生可能性。
	適合所有道路使用者的速限	依據道路幾何形狀、車道密度及歷史事故等，設定道路限速，避免駕駛員於不清楚道路環境情形，以自身合適車速行駛，導致事故發生。
車道	路面邊緣加寬	路面邊緣寬度從 4 英吋（約 10 公分）最小線寬增加至 6 英吋（）最大線寬時，可提升駕駛對行車邊緣的感知，進而降低車道偏離事故風險。
	增強水平曲線輪廓	透過增強輪廓，提醒駕駛注意即將到來的彎道及其方向和銳度，以降低視覺陷阱及提早應變行駛操作。
	隆起標線	運用行經隆起標線發出的震動及聲音提醒駕駛車輛已偏離車道。
	緩斜路緣	距離路面橫坡約 30 度位置設置緩斜路緣，消除路面邊緣垂直下降、因侵蝕和輪胎磨損暴露的可能性，進而降低邊緣落差事故可能性及嚴重度。
	彎道路邊設計改進	於彎道外側設置淨空區、拓寬路肩或邊坡平坦，使駕駛於偏離車道時，可更容易穩定及重新控制車輛，並降低事故嚴重程度。
	中央安全護欄	運用縱向護欄分隔對立的交通流，以及阻隔單向事故撞擊後，跨越中間的可能性

參考資料：FHWA Making Our Roads Safer

三、研究方法

從文獻回顧中得知目前臺灣在交通安全管理作業現況與近年實證研究成果，揭示我國汽車運輸業在安全管理制度設計與實務執行間，仍存在明顯落差。其中，高風險族群多集中於駕駛管理鬆散、制度化程度不足與作業特性複雜之業者，顯示亟需建立具科學依據且能應用於查核與輔導之風險辨識工具。為因應此一需求，本處將進一步進行資料整理與量化指標建構，奠定後續風險分級與優先輔導機制的基礎。

3.1 研究目的及流程

透過高公局中分局轄區之 110-113 年公警全般事故資料，與中分局道路環境變數共 24 種，分析目的欲了解哪些環境變數組合下會容易造成一般車輛的危險（風險），將道路基礎環境及全般事故切分為每公里，亦即組合成每公里路段上每年發生之事故數量及其環境因子進行分析。

考量道路環境變數之間存在共變關係，故優先將 24 種環境變數依性質及重要程度分類為道路幾何、道路環境、交通管制設施、車流量等 4 大類別，並依類別依序使用逐步迴歸分析（stepwise regression analysis）找尋顯著變數。另文獻雖皆以負二項或卜瓦松迴歸進行分析，惟本文希冀透過環境變數對交通事故數的「方向與相對影響力」，以更直觀方式進行篩選，而非精準預測事故數，因此以 OLS 模型為依據，依環境變數之順序逐步投入挑選變數。將顯著變數投入隨機森林（Random Forest, RF）及極限梯度提升（XGboost）方法進行特徵重要性分析，並結合顯著變數之係數進行解釋。

3.2 研究方法

（一）線性迴歸（Linear Regression, LR）

多元線性迴歸（Multiple Linear Regression），對應的數學表達式為：

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \epsilon \quad (1)$$

其因自變數數量增加，可能產生多重共線性（Multicollinearity）問題，導致迴歸係數不穩定並影響模型解釋性。常見的解決方案包括 Lasso 迴歸（L1 正則化，用於變數選擇）、逐步迴歸（根據統計顯著性增減變數）以及方差膨脹因子（VIF）檢測等。為了估計最佳的迴歸係數，通常使用最小平方方法（Ordinary Least Squares, OLS），其估計公式如下，其中 X 是自變數矩陣；是 Y 應變數向量：

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (2)$$

（二）決策樹（Decision tree, DT）

決策樹是一種基於示例學習的模型，透過遞歸劃分數據集來建構樹狀結構，使其能夠進行分類或迴歸預測。在分類問題中，決策樹使用資訊增益來選擇最佳特徵，該方法依賴於熵（Entropy）來衡量數據的不確定性：

$$H(D) = -\sum_{i=1}^c p_i \log_2 p_i \quad (3)$$

其中 p_i 為某類別在數據集中的比例，當所有數據屬於單一類別時，其熵為 0；當數據均勻分布於不同類別時，其熵為最大。資訊增益用於衡量某特徵對於降低熵的貢獻，定義如下：

$$Gain(A) = H(D) - \sum_{i=1}^v \frac{|D_i|}{|D|} H(D_i) \quad (4)$$

其中 D_i 是屬性劃分後的子數據集，當某屬性的劃分能最大程度減少數據的不確定性，則選擇該屬性作為分裂節點。在迴歸問題中，決策樹無法使用熵，而是透過均方誤差（MSE）來衡量預測的準確性，計算方式如下：

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2 \quad (5)$$

其中 y_i 是觀測值； $\bar{y}$ 為該節點內樣本的均值。決策樹會遍歷所有可能的分裂點，選擇能最大程度降低 MSE 的特徵。由於決策樹容易過度擬合，剪枝（Pruning）技術被用來控制模型複雜度，提升泛化能力。預剪枝透過設定閾值來決定何時停止劃分，而後剪枝則在樹建構完成後，移除對準確度影響不大的節點。除了過擬合問題，決策樹在特徵選擇時也可能偏好取值較多的屬性，因為它們通常帶來較高的資訊增益。為了解決這一問題，引入增益比（Gain Ratio）來正規化資訊增益，公式如下：

$$Gain Ratio(A) = \frac{Gain(A)}{IV(A)} \quad (6)$$

其中 $IV(A)$ 衡量屬性 A 本身的內在不確定性：

$$IV(A) = -\sum_{i=1}^v \frac{|D_i|}{|D|} \log_2 \frac{|D_i|}{|D|} \quad (7)$$

這一調整避免決策樹過度依賴某些取值範圍較大的特徵，使其選擇更具代表性的分裂節點。這些技術確保了決策樹在機器學習領域的廣泛應用，使其能夠處理高維數據與非線性關係，並在許多實際場景中被廣泛使用。

（三）隨機森林（Random Forest, RF）

它是透過構建多棵決策樹來提升分類與迴歸的準確性，並降低過擬合風險。其基礎來自 Bagging（Bootstrap Aggregating），每棵決策樹皆基於隨機抽樣的數據集與隨機選擇的特徵進行訓練，使得決策樹之間的相關性降低，進而提高模型的泛化能力。在分類問題中，隨機森林透過多數決來確定最終類別，即每棵決策樹進行預測，並選擇得票數最多的類別作為輸出：

$$\hat{Y} = \arg \max_j \sum_{k=1}^K I(h_k(X) = j) \quad (8)$$

其中 $I(\cdot)$ 為指示函數。隨機森林的泛化誤差由決策樹的強度（Strength）和相關性（Correlation）影響，其上界可表示為：

$$PE^* \leq \frac{\bar{\rho}(1-s^2)}{s^2} \quad (9)$$

其中 s 為平均 margin， $\bar{\rho}$ 則表示決策樹之間的相關性。當決策樹的 margin 增加、相關性降低時，模型的泛化誤差會減少。而在迴歸問題中，隨機森林的預測值來自所有決策樹的平均：

$$\hat{Y} = \frac{1}{K} \sum_{k=1}^K h_k(X) \quad (10)$$

此方法可視為降噪技術，能夠降低模型變異性，提高預測穩定性。其泛化誤差由均方誤差（MSE）衡量，而誤差上界為：

$$MSE_{RF} \leq \bar{\rho} \cdot MSE_{Single Tree} \quad (11)$$

（四）極限梯度提升（eXtreme Gradient Boosting, XGboost）

XGBoost 是一種基於梯度提升決策樹（Gradient Boosting Decision Tree, GBDT）的機器學習模型，透過逐步優化來提升模型的預測效能。為了確保模型在學習過程中保持穩定，XGBoost 採用了梯度提升策略，透過計算損失函數的一階與二階導數來調整樹的權重，進而改善模型的學習能力並加速收斂。

XGBoost 的目標函數由兩部分組成，分別是損失函數（Loss Function）與正則化項（Regularization Term）。損失函數用於衡量模型的預測結果與實際值之間的差異，例如對於迴歸問題，常見的損失函數包括均方誤差（Mean Squared Error, MSE）：

$$L = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (12)$$

而在分類問題中，則通常使用交叉熵損失函數（Cross-Entropy Loss）來衡量預測概率與真實標籤之間的誤差：

$$L = -\sum_{i=1}^n y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i) \quad (13)$$

為了進一步提升模型的泛化能力，XGBoost 在目標函數中引入了正則化項來控制模型的複雜度，主要透過 L1 正則化（Lasso）與 L2 正則化（Ridge）來約束決策樹的參數，避免模型過擬合：

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (14)$$

其中 T 代表樹的葉節點數； w_j 是葉節點的權重； γ 負責控制樹的複雜度； λ 而是 L2 正則化係數。XGBoost 透過梯度提升（Gradient Boosting）來優化損失函數，並使用二階泰勒展開（Taylor Expansion）來近似損失函數，以提高計算效率。每次迭代時，XGBoost 會計算損失函數的一階導數（梯度）與二階導數（曲率），並根據這些資訊來決定如何調整模型參數：

$$g_i = \frac{\partial L}{\partial \hat{y}_i}, \quad h_i = \frac{\partial^2 L}{\partial \hat{y}_i^2} \quad (15)$$

其中 g_i 代表梯度，決定了損失函數的下降方向，而 h_i 表示曲率，幫助調整步長，以避免更新過快或過慢。此外，XGBoost 透過最大化增益（Gain）來決定最佳的分裂節點，增益計算如下：

$$Gain = \frac{1}{2} \left[\frac{(\sum g_L)^2}{\sum h_L + \lambda} + \frac{(\sum g_R)^2}{\sum h_R + \lambda} - \frac{(\sum g_P)^2}{\sum h_P + \lambda} \right] - \gamma \quad (16)$$

其中 g_L 和 g_R 為左右子節點的梯度總和； h_L 和 h_R 為相應的二階導數總和； λ 為正則化參數；而 γ 則負責抑制不必要的節點增長，確保模型的簡潔性。透過這樣的機制，XGBoost 在保持高效計算的同時，也能提供穩定且準確的預測結果，使其在各種應用場景中表現優異。

(五) SHAP (SHapley Additive exPlanations)

SHAP 是一種基於 Shapley 值的模型解釋方法，透過計算各特徵對預測結果的邊際貢獻，來提供公平且透明的特徵影響分析。SHAP 透過計算不同特徵組合的貢獻值，來評估每個特徵對模型輸出的影響，確保解釋方式具備公平性與一致性。對於 Shapley 值，其計算方式如下：

$$\Phi_j = \sum_{S \subseteq N \setminus \{j\}} \frac{|S|!(|N|-|S|-1)!}{|N|!} [f(S \cup \{j\}) - f(S)] \quad (17)$$

其中 Φ_j 代表特徵 j 的 SHAP 值； N 是所有特徵的集合； S 為不包含 j 的特徵子集； $f(S)$ 是僅考慮特徵集合 S 的模型輸出；而 $f(S \cup \{j\}) - f(S)$ 則表示加入特徵 j 後對預測結果的影響。透過遍歷所有可能的特徵組合，並計算其對模型輸出的邊際影響，SHAP 能夠準確衡量各特徵的貢獻。

四、資料分析

4.1 資料蒐集

本文中分局道路環境變數資料中，係以每百公尺為一筆資料，每筆資料中共有 24 種環境變數，另交通事故資料為高公局中分局轄區之 110-113 年公警全般事故資料。為進行分析之簡化，將每百公尺的資料整併為每公里，其中路段共計 748 段，並有將國 4 未開通時間路段資料移除不納入分析（移除 110 年國道四號 26K~28K 及 110 年至 111 年國道四號 17K~26K）；而 Y 變數為 110 年至 113 年全般事故件數，其資料合併後共有 2952 筆資料，資料結構如圖 3。

里程	環境變數1	...	環境變數N	全般事故件數
100				1
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
195				0

X變數
Y變數

圖 3 資料結構範例

為避免車流量過度配適（overfitting）模型，依據本文分析重點，將 24 種環境變數分為四大種項目類別，包含道路幾何線型、環境變數、交通管制措施、車流量，並作為以下說明變數基本統計及後續投入分析變數之順序依據，環境變數投入資料順序與說明彙整如下表 4。

表 4 環境變數投入順序與變項名稱彙整表

變數項目類別	投入順序	投入變數	變數類型	變項名稱或計算方式
道路幾何線型	1	彎道半徑 R 值	連續	取該公里（每公里裡面有 10 個百公尺路段的數值）最小值，如 A 值為空值則取 6000、R 值空值或大於 8000 者即用 8000 取代
	2	彎道半徑 A 值	連續	將數值透過標準化 ($R_{new} = \frac{R_{min}-R_{old}}{R_{min}-R_{max}}$, $R_{min} = 1000$, $R_{max} = 8000$) 後投入分析，數值介於 0 至 1 之間，越大代表直線
	3	縱向坡度_上坡	連續	縱向上下坡互斥概念，分別帶入
	4	縱向坡度_下坡	連續	縱向上下坡互斥概念，分別帶入
	5	橫向坡度（絕對值）	連續	如該變數顯著則 6、7 不投入
	6	橫向坡度_上坡	連續	橫向上下坡互斥概念，分別帶入
	7	橫向坡度_下坡	連續	橫向上下坡互斥概念，分別帶入
	8	車道數	類別	1) 車道數_2；2) 車道數_3；3) 車道數_4；4) 車道數_3 轉 4 生成 4 個虛擬變數（dummy variable），以車道數_2 為參照組，投入其他 3 個自變項至迴歸方程式
	9	加速車道_有	二元	
	10	減速車道_有	二元	
	11	路肩（有否 3m）_有	二元	
	12	避車彎_有	二元	
環境變數	13	逆光路段_有	二元	
	14	照明路段	類別	1) 否；2) 是；3) 混合
	15	橋梁_是	二元	
	16	隧道_是	二元	
	17	隔音設施_是	二元	
交通管制措施	18	常態開放路肩_有	二元	
	19	E 牌	連續	該公里（每公里裡面有 10 個百公尺路段的數值）數值累加
	20	T 牌	連續	
	21	LED 牌面	連續	
	22	定點測速_有	二元	
	23	速限	類別	1) 80KPH；2) 100KPH；3) 110KPH
車流量	24	車流量	連續	計算各路段（每公里）每一年的總車流量 資料類型：每分鐘 VD 流量（高公局 open data VD 動態資訊） 資料時間：110 年 1 月 1 日 ~113 年 12 月 31 日

另 Y 變數事件數，以公警全般事故資料 A1 至 A3 事件數，並以每年每公里作為統計依據，資料統計表如表 5。

表 5 每年每公里事故件數 (Y 變數) 之資料筆數統計表

Y 變數 - 事故件數	資料筆數	資料比例	Y 變數 - 事故件數	資料筆數	資料比例	Y 變數 - 事故件數	資料筆數	資料比例
0	155	5.2%	10	122	4.1%	20	30	1.0%
1	128	4.3%	11	99	3.3%	21	31	1.0%
2	273	9.2%	12	86	2.9%	22	27	0.9%
3	200	6.7%	13	77	2.6%	23	21	0.7%
4	233	7.8%	14	65	2.2%	24	19	0.6%
5	208	7.0%	15	59	2.0%	25	23	0.8%
6	196	6.6%	16	54	1.8%	26	20	0.7%
7	172	5.8%	17	43	1.4%	27	16	0.5%
8	161	5.4%	18	35	1.2%	28	20	0.7%
9	143	4.8%	19	37	1.2%	29 以上	219	7.4%

4.2 逐步迴歸篩選變數

依照前述變數投入順序，依序確認該變數是否顯著，如顯著則保留，不顯著則刪除，其中圖 4 以第一個變數 - 彎道曲率半徑 R 值為例，P 值為顯著，則保留該變數；投入第二順位變數 - 彎道曲率半徑 A 值（克羅索曲線）並保留前述顯著變數 - 彎道曲率半徑 R 值，其 P 值皆為顯著，則保留兩個變數，依此類推依序投入所有環境變數。

Step 1：彎道曲率半徑 R 值

OLS Regression Results						
Dep. Variable:	件數	R-squared:	0.068			
Model:	OLS	Adj. R-squared:	0.068			
Method:	Least Squares	F-statistic:	216.5			
Date:	Wed, 19 Mar 2025	Prob (F-statistic):	2.44e-47			
Time:	11:36:22	Log-Likelihood:	-11358.			
No. Observations:	2952	AIC:	2.272e+04			
DF Residuals:	2950	BIC:	2.273e+04			
DF Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	6.6561	0.335	19.868	0.000	5.999	7.313
彎道曲率半徑 R 值	15.6383	1.063	14.713	0.000	13.554	17.722
Omnibus:	1876.133	Durbin-Watson:	0.823			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	24135.582			
Skew:	2.855	Prob(JB):	0.00			
Kurtosis:	15.792	Cond. No.	5.41			

Step 2：彎道曲率半徑 A 值(克羅索曲線)

OLS Regression Results						
Dep. Variable:	件數	R-squared:	0.073			
Model:	OLS	Adj. R-squared:	0.072			
Method:	Least Squares	F-statistic:	115.7			
Date:	Wed, 19 Mar 2025	Prob (F-statistic):	4.38e-49			
Time:	11:35:35	Log-Likelihood:	-11351.			
No. Observations:	2952	AIC:	2.271e+04			
Df Residuals:	2949	BIC:	2.273e+04			
Df Model:	2					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	6.2646	0.350	17.880	0.000	5.638	6.952
彎道曲率半徑 R 值	13.6454	1.188	11.490	0.000	11.317	15.974
彎道曲率半徑 A 值(克羅索曲線)	2.3994	0.643	3.730	0.000	1.138	3.661
Omnibus:	1870.595	Durbin-Watson:	0.832			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	23915.372			
Skew:	2.846	Prob(JB):	0.00			
Kurtosis:	15.730	Cond. No.	6.58			

圖 4 逐步迴歸篩選變數步驟範例

前述流程操作完成後，顯著變數共計 11 個，作為後續投入隨機森林及 XGBoost 模型進行重要程度分析之變數，其分析結果如下表 6。

依分析結果彙整表顯示，道路線型部分，R 值或 A 值越高（亦表示道路越直），影響事故件數越多。其結果係因高公局中分局轄區中，臺中都會區路段、彰化路段皆是 R 值較大的路段，惟路段因車流量大，且固定時段會出現重現性壅塞，因此事故量較大，其道路幾何條件結果才跟過往先驗知識和文獻回顧之結果較為不同，惟其係數很小，影響程度小。

道路環境部分，逆光路段、常態開放路肩、車道數越多、有加減速車道之路段、定點測速路段等，影響事故件數越多。另車流量部分，投入後會導致車道數之變數不顯著，表示車流量及車道數兩者變數存在共變性，因此不將車流量納入後續機器學習的變數之一，也避免模型過度受到車流量之影響。

表 6 逐步迴歸分析結果彙整表

變數	係數	標準差	t 值	p 值
彎道半徑 R 值	0.0007	0.0001	9.089	<0.001
彎道半徑 A 值	0.0004	0.0001	3.726	<0.001
橫向坡度（絕對值）	0.5679	0.175	3.237	0.001
車道數				
車道數_2	6.5669	0.769	8.541	<0.001
車道數_3	8.6733	0.908	9.555	<0.001
車道數_4	10.9669	1.037	10.576	<0.001
加速車道_有	1.5152	0.491	3.083	0.002
減速車道_有	1.2496	0.491	2.545	0.011
逆光路段_是	29.9940	3.115	9.628	<0.001
橋梁_是	-2.0732	0.427	-4.853	<0.001
隔音設施_有	2.3318	0.418	5.581	<0.001
常態開放路肩_有	8.6176	1.331	6.474	<0.001
定點測速_有	2.1406	0.716	2.989	0.003
常數項	-0.9315	0.798	-1.168	0.243
R-square	0.175			
Adj. R-square	0.171			
F 統計量	47.77(p<0.001)			
AIC	22390			
樣本數	2938			

4.3 機器學習分析成果

將前述逐步迴歸分析之顯著變數，分別投入 RF 及 XGB 模型，用以了解其特徵重要性，並進一步查看特徵重要性排名順序是否一致，其模型特徵值及排序如下表 7，兩者預測之值分別是 0.6821 及 0.6599，一般大於 0.6 即算該模型具有解釋力。

表 7 RF 及 XGB 模型結果彙整表

變數	RF-Mean Decrease in Impurity	RF 特徵排名	XGB-F score	XGB 特徵排名
彎道半徑 R 值	0.2735	1	965	1
彎道半徑 A 值	0.1834	2	887	2
橫向坡度（絕對值）	0.1523	3	606	3
車道數				
車道數_2	0.02	12	158	8
車道數_3	0.0159	13	94	10
車道數_4	0.0244	11	61	11
加速車道_有	0.0447	7	176	7
減速車道_有	0.0707	4	210	6
逆光路段_是	0.0453	6	31	13
橋梁_是	0.0445	8	256	4

變數	RF-Mean Decrease in Impurity	RF 特徵排名	XGB-F score	XGB 特徵排名
隔音設施_有	0.0566	5	231	5
常態開放路肩_有	0.0333	10	39	12
定點測速_有	0.0335	9	108	9
R-square	0.6821		0.6599	
樣本數	2938			

另為了解每個特徵對最終預測值的貢獻是多少，並能夠進一步解釋特徵對預測值的正向或負向貢獻，以及交互作用的影響，本文透過 RF 及 XGB 模型計算 SHAP 值，並繪製蜂巢圖，如圖 5 及圖 6。其中，利用 RF 及 XGB 模型計算 SHAP 值之特徵重要性排名大致相符，且兩者模型所分析出的特徵值，其正負相關性大致相符。

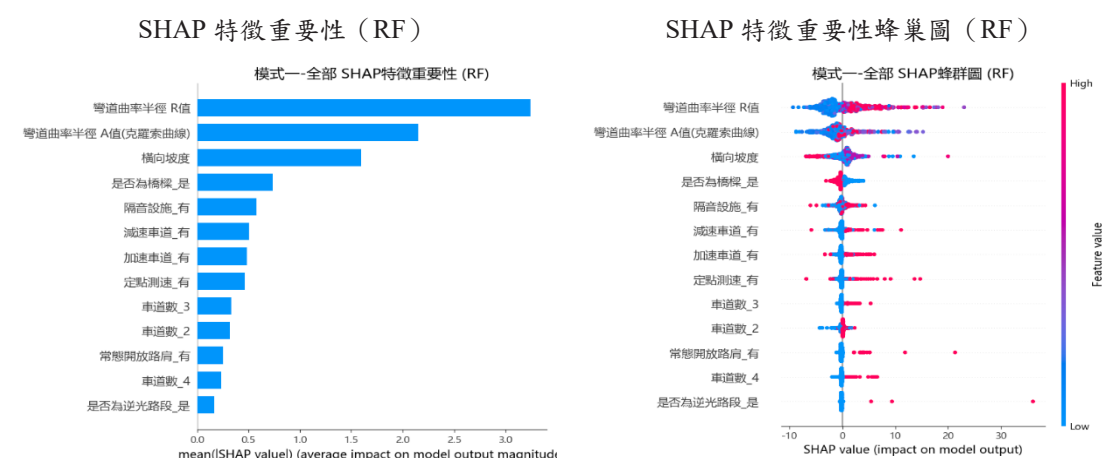


圖 5 SHAP 特徵重要性及蜂巢圖 (RF)

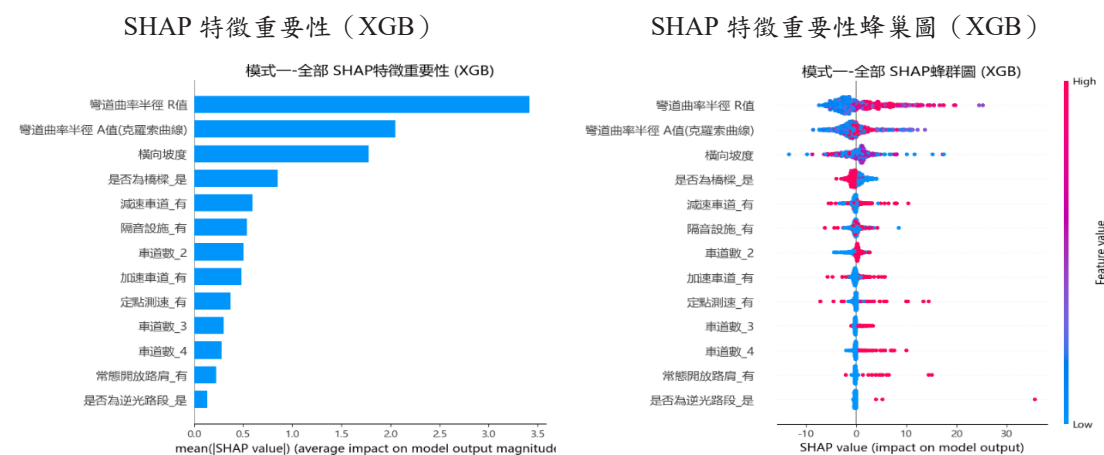


圖 6 SHAP 特徵重要性及蜂巢圖 (XGB)

另比對表 6 逐步迴歸分析結果彙整表中各變數之係數，顯示彎道半徑 R 值、橋樑_是、隔音設施_有、加速車道_有、減速車道_有、車道數、常態開放路肩_有、定點測速_有、逆光路段_是等 9 個變數中，其分析結果和 RF 及 XGB 模型計算 SHAP 值之正負相關性相符。

彙整 RF 及 XGB 模型特徵值排名及 SHAP 值排名，並不考慮逐步迴歸分析與機器學習分析之特徵值（係數）正負結果相異之變數，其中前五名重要之變數分別是彎道半徑 R 值、減速車道_有、隔音設施_是、橋梁_是及加速車道_有，如表 8。

表 8 RF 及 XGB 模型特徵值 SHAP 值排名彙整表

逐步迴歸之顯著變數	迴歸是否與 SHAP 值正負性相符	模型特徵重要性		SHAP 特徵重要性		平均
		RF	XGB	RF	XGB	
彎道半徑 R 值	是	1	1	1	1	1.0
彎道半徑 A 值（克羅索曲線）		2	2	2	2	2.0
橫向坡度		3	3	3	3	3.0
減速車道_有	是	4	4	6	4	4.5
隔音設施_有	是	5	6	5	5	5.3
橋梁_是	是	8	5	4	6	5.8
加速車道_有	是	7	7	7	8	7.3
定點測速_有	是	9	8	8	7	8.0
逆光路段_是	是	6	9	13	9	9.3
常態開放路肩_有	是	10	10	11	10	10.3
車道數_2	是	12	11	10	11	11.0
車道數_3	是	13	12	9	13	11.8
車道數_4	是	11	13	12	12	12.0

4.4 小結

綜合分析結果，彎道半徑 R 值較大、有減速車道之路段、有隔音設施之路段以及有加速車道之路段，其對於高公局中分局轄區路段之事故件數為正相關。橋梁路段，其對於高公局中分局轄區路段之事故件數為負相關。另分析其變數之正負相關可能原因如表 9。

表 9 本文各模型前五名重要變數之原因分析彙整表

各模型前五名重要變數	逐步迴歸模型之係數	RF 及 XGB 模型特徵重要性平均排名	原因分析
彎道半徑 R 值	0.0007	1.0	R 值較大的路段多在臺中都會區路段及彰化路段，其路段車流相較其他路段多，事故量較大，因此 R 值為正相關，惟影響程度小。
減速車道_有	1.2496	4.5	有減速車道之路段，因車流於此路段之車速變異性大，常有出口壅塞或有減速之情況，因此其容易造成追撞事故。
隔音設施_有	2.3318	5.3	有隔音設施之路段，推測係因彎道或開放路肩，其隔音設施對於用路人之視距影響，造成其分析結果。

各模型前五名重要變數	逐步迴歸模型之係數	RF 及 XGB 模型特徵重要性平均排名	原因分析
橋梁_是	-2.0732	5.8	橋梁路段分析結果為負相關，推測因橋梁路段於國道 3 號、國道 4 號及國道 6 號之路段較多，國道 1 號多為路堤及路塹路段。 又於高公局中分局轄區中，因國道 3 號、國道 4 號及國道 6 號車流量較小，其事故量總合約為全轄區之 38.4%；國道 1 號事故量約為全轄區之 61.6%，因此分析結果與其轄區資料特性有很大的相關性。
加速車道_有	1.5152	7.3	有加速車道之路段，因車流於此路段之常有匯入之車流，易於主線上游車流產生衝擊波，因此其容易造成匯入之同向擦撞及上游車流之追撞事故。

五、交通工程加強作為實作

經由前節分析結果及原因分析，本文針對事故較多，且篩選出彎道半徑 R 值較大、有加減速車道之路段，進行交通工程加強作為之實作，以期透過分析結果，實際落地於道路環境之策進作為。

1. 彎道半徑 R 值較大：國 1 北向北斗到員林（220k-212k）

以此路段為例，彎道半徑 R 值較大，速度較高。又因其周日重現性壅塞，事故多為追撞事故。本文透過國 1 北 217k 上游處增設「保持安全間距」之牌面，國 1 北 216k 跨越橋下方增設「行車安全距離辨識標線」及牌面進行加強，詳如圖 7。



圖 7 彎道半徑 R 值較大之交通工程加強作為

2. 減速車道路段：國 3 南向和美到彰化系統（192k-196k）

以此路段為例，因減速車道銜接出口後，便會銜接國道 1 號，惟因國 1 主線壅塞，造成車速變異性大，事故多為追撞事故。本文透過減速車道上游處之國 3 北 195.3k、國 3 北 195.1k 及國 3 北 193.6k 的外側車道，增設「請注意車前動態」牌面進行加強，詳如圖 8。

國 3 北 193.6k



國 3 北 195.1k



國 3 北 195.3k



圖 8 速車道路段之交通工程加強作為

3. 加速車道路段：國 1 北向豐原至臺中系統（168k-165k）

以此路段為例，加速車道雙白線約 60 公尺後，以穿越虛線方式讓車輛匯入主線，惟車輛可能尚未加速到主線速限，易於主線上游車流產生衝擊波，也易因匯入車速不夠，容易產生同向擦撞及上游車流之追撞事故。本文透過延長雙白線 180 公尺，提供車輛加速至速限才得以匯入主線之長度，詳如圖 9。

延伸雙白線前



延伸雙白線後



參考資料：google map

圖 9 加速車道路段之交通工程加強作為

六、結論

本文透過逐步迴歸、XGB、隨機森林等機器學習方法進行分析道路幾何及環境變數對於事故件數的影響，並得到下列結論：

1. 彎道半徑 R 值較大、有減速車道、有隔音設施、有加速車道、有定點測速、有逆光路段、有常態開放路肩，以及車道數越多之路段，其對於高公局中分局轄區路段之事故件數為正相關。橋梁路段，其對於高公局中分局轄區路段之事故件數為負相關。
2. 高公局中分局轄區路段中， R 值較大的路段因位於車流量較大之都會區，又橋梁段多位於國 3、國 4 及國 6 等車流量較低之路線，因此道路幾何及環境變數分析結果，才與先驗知識及文獻回顧之結果較為不同。

本文乃針對國道中區路段進行道路幾何及事故件數分析，據以找出關鍵路段或條件，以研擬加強作為，但國道各路段之交通特性不盡相同，未來應用時應適時依交通工程司專業判斷予以調整及分析。

七、參考文獻

- iRAP Star Rating and Investment Plan Manual(2024), International Road Assessment Programme (iRAP)
- CAREC Road Safety Engineering Manual 5(2022), Asian Development Bank
- Venkataraman Shankar, Fred Mannering, Woodrow Barfield(1995). Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. Accident Analysis & Prevention, Volume 27, Issue 3, 371-389
- Kopelias, P., Papadimitriou, F., Papandreou, K., & Prevedouros, P. (2007). Urban Freeway Crash Analysis: Geometric, Operational, and Weather Effects on Crash Number and Severity. Transportation Research Record, 2015(1), 123-131
- Chan, Chun & Huang, Baoshan & Richards, Stephen. (2009). Relationship Between Highway Pavement Condition, Crash Frequency, and Crash Type. Journal of Transportation Safety & Security. 1. 268-281.
- Xiaofei Wang, Siyu Li, Xinwei Li, Yinhai Wang, Qiang Zeng (2023). Effects of geometric attributes of horizontal and sag vertical curve combinations on freeway crash frequency. Accident Analysis & Prevention, Volume 186, 107056
- Xiao Wen, Yuanchang Xie, Liming Jiang, Yan Li, Tingjian Ge(2022). On the interpretability of machine learning methods in crash frequency modeling and crash modification factor development. Accident Analysis & Prevention, Volume 168, 106617
- J.P.S. Shashiprabha Madushani, R.M. Kelum Sandamal, D.P.P. Meddage, H.R. Pasindu, P.I. Ayantha Gomes(2023). Evaluating expressway traffic crash severity by using logistic regression and explainable & supervised machine learning classifiers. Transportation Engineering, Volume 13, 100190
- Yang Yang, Kun Wang, Zhenzhou Yuan, Dan Liu(2022). Predicting Freeway Traffic Crash Severity Using XGBoost-Bayesian Network Model with Consideration of Features Interaction. Journal of Advanced Transportation, Volume 2022.
- Wang, J., Ma, S., Jiao, P., Ji, L., Sun, X., & Lu, H. (2023). Analyzing the Risk Factors of Traffic Accident Severity Using a Combination of Random Forest and Association Rules. Applied Sciences, 13(14), 8559.
- Hollnagel, E., N  bo, A., & Lau, I. V. (2005). A systemic model for driver-in-control. 2. Making Our Roads Safer | One Countermeasure at a Time(2021), Federal Highway Administration Office of Safety