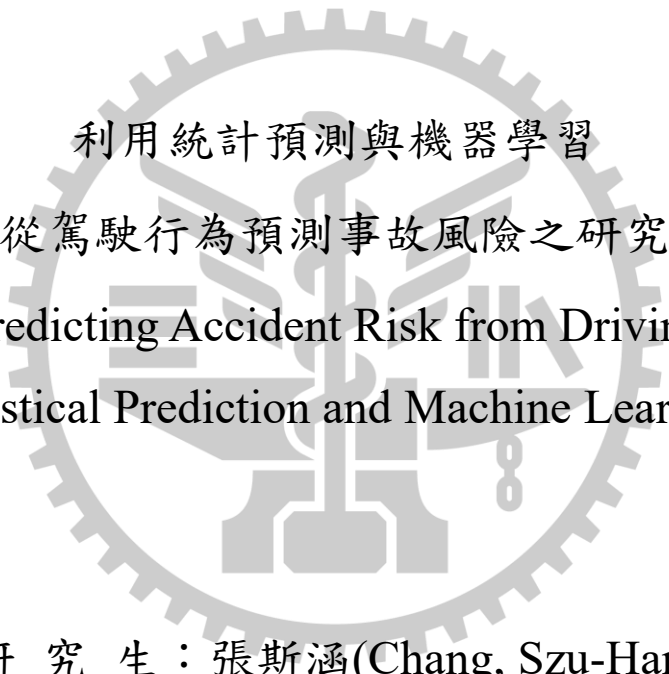


國立陽明交通大學
運輸與物流管理學系
碩士論文

Department of Transportation and Logistics Management
National Yang Ming Chiao Tung University
Master Thesis



利用統計預測與機器學習
從駕駛行為預測事故風險之研究
Research on Predicting Accident Risk from Driving Behavior by
Statistical Prediction and Machine Learning

研 究 生：張斯涵(Chang, Szu-Han)

指導教授：王晉元(Wang, Jin-Yuan)

中華民國一一二年九月

September 2023

利用統計預測與機器學習

從駕駛行為預測事故風險之研究

Research on Predicting Accident Risk from Driving Behavior
by Statistical Prediction and Machine Learning

研究生：張斯涵

Student：Szu-Han Chang

指導教授：王晉元 博士

Advisor：Dr. Jin-Yuan Wang

國立陽明交通大學

運輸與物流管理學系

碩士論文

A Thesis

Submitted to Department of Transportation and Logistics Management
College of Management
National Yang Ming Chiao Tung University
in Partial Fulfillment of the Requirements
for the Degree of
Master of Science
in
Traffic and Transportation

September 2023

Taiwan, Republic of China

中華民國一十二年九月

誌謝

研究所生活和寫論文過程實在得到太多支持與幫助，有太多值得感謝的人事物。常常覺得研究所的日子很開心、很充實，甚至很有意義、很有成就感，雖然也有苦澀的部分，但好像過了就忘了，我想是家人、老師、朋友以及許多人事物帶給我的養分，滋養了我的研究所生活，讓挑戰都迎刃而解。獲得這些養分，並非理所當然，感謝我的家人、老師、朋友以及許多其他人事物。

能順利完成研究所學業，要感謝我的家人。謝謝我的爸媽，滿足我的食衣住行育樂，並給我溫暖、自由、陪伴、支持，可以說是給了我一切，我常常覺得幸福快樂，非常感謝。謝謝我的弟弟，每天跟我鬥嘴，希望你也覺得有趣，也謝謝你的好脾氣。謝謝我的阿婆、姑姑、姑丈、阿爸、阿姨、哥哥、姐姐，帶給我許多人生啟發，讓我看見人性樂觀、寬容、堅韌的一面，很感謝有這樣的家族。想讀研究所、進入研究所、讀完研究所，並非人生的必經道路，很感謝有家人的支持與啟發，我才有機會順利完成研究所學業。

能順利完成碩士論文，要感謝我的老師。尤其是我的指導教授，王晉元老師。王老師真的是很用心、很負責的老師，感謝王老師總是很願意花時間和學生討論、很仔細提供論文建議、很快速回應學生。同時因為跟老師做計畫，才熟練了數據處理能力，不論是數值、文字，甚至是地理資訊，都能處理，完全補足了課程較少涵蓋到，但很實用的部分，非常幸運我的指導教授是王老師，收穫已超出我的預期。若沒有王老師，我想我很難順利完成碩士論文。還要感謝研究所的老師，尤其是交通安全領域的老師，林陳佑老師、吳昆峰老師、吳宗修老師，讓我體會交通安全研究的有趣與價值，以及傳授實用的分析工具，我才能順利完成交通安全相關的研究。也要感謝大學的老師，戴佐敏老師、魏健宏老師、沈宗緯老師、林珮琄老師、陳文字老師、李威勳老師等等，讓我開始了解知識的迷人之處，在這之前我並不懂得欣賞知識的美好。不論是交大運管還是成大交管的老師，傳授的知識不只讓我順利完成碩士論文，更改變了我學習及思考的態度。最後要感謝昱光學長提供資料，像老師一樣為我解答了許多疑問，謝謝。

能順利度過求學生活，要感謝我的朋友。感謝研究室同學書豪、欣妤、凱韻能一起調皮搗蛋、互相幫忙、度過危機，你們真的是很酷的人，友善、隨和、獨立、天然、純粹、脫線，我很喜歡。感謝佩恩常常陪玩、陪讀，甚至幫我看論文的文句，你也是很酷的人，善良、可愛、體貼、搞笑、冒失，我很喜歡。感謝秀蔭姊、品蓁學姊、日隆、語芳、子賢學長姊，以及蘇旆、雅絮、宣宇、子傑碩一親故，讓王 lab 運轉得十分順利，感謝有你們。還有謝謝研究所同學、阿澤、臭三八、高中同學、國中麻吉，謝謝你們的友善、幽默、真誠，讓我的求學生活過得很愉快。

張斯涵

2023 年 9 月於交大

利用統計預測與機器學習從駕駛行為預測事故風險之研究

研究生：張斯涵

指導教授：王晉元

國立陽明交通大學運輸與物流管理學系碩士班

摘 要

事故風險預測能辨識出高風險駕駛以及高風險因子，進而幫助主管機關，如政府部門、客運業者，了解應加強管理的對象以及構思管理的方案，以預防事故發生。然過往相關研究，常面臨事故資料蒐集時間長、變數眾多導致變數影響不顯著之困境。為了解決以上困境，本研究欲提出完整之事故風險預測方法，內容分為事故風險衡量指標快速確立、事故風險影響因子取用及挑選、事故風險預測模型建立及挑選，三大部分。事故風險衡量指標快速確立之方法為，輸入 ADAS 警示次數進入 K-means，劃分事故風險群，並以群間平均數比較與群間平均數差異檢定之結果，作為命名事故風險群參考。事故風險影響因子取用及挑選之方法為，取出駕駛行為特徵及環境因子後，以 t 檢定、RFE、Lasso 挑選出重要的變數。事故風險預測模型建立及挑選之方法為，訓練羅吉斯迴歸、判別分析、決策樹、SVM 後，以平均 AUC 挑選出最佳模型。本研究以雙北地區大客車 ADAS 警示及車速車距為資料，實際測試以上分析方法，結果如下：駕駛分為相對低風險群以及相對高風險群，且兩群之 ADAS 警示次數有顯著不同，故兩群的事故風險有顯著差異。三種變數挑選方法的結果並不完全相同，若只參考單一方法恐怕有些偏頗，應綜合考量三種方法。最終模型為羅吉斯迴歸，因其平均 AUC 最大。

關鍵詞：事故風險、駕駛行為、變數選擇、統計預測、機器學習

Research on Predicting Accident Risk from Driving Behavior by Statistical Prediction and Machine Learning

Student : Szu-Han, Chang

Advisor : Dr. Jin-Yuan Wang

Department of Transportation and Logistics Management

National Yang Ming Chiao Tung University

Abstract

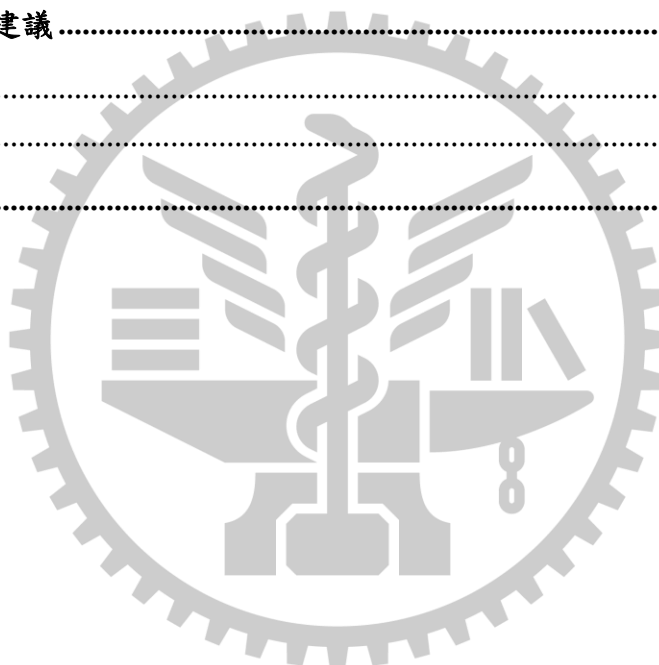
The prediction of accident risks enables the identification of high-risk drivers and high-risk factors. However, previous related studies often encountered challenges in the length of collecting accident data and insignificance when including too many variables. To address these challenges, this study aims to propose a comprehensive accident risk prediction method, which can be divided into three parts: rapid establishment of accident risk indicators, extraction and selection of risk influencing factors, and establishment and selection of accident risk prediction models. The establishment method for accident risk indicators involves inputting ADAS warnings into K-means to cluster accident risk groups. The results of comparing group means and group means difference tests are used to name the accident risk groups. The method for extracting and selecting risk influencing factors involves extracting driver behavior characteristics and environmental factors, and then using t-tests, RFE, and Lasso to identify important variables. The method for establishing and selecting accident risk prediction models involves training logistic regression, discriminant analysis, decision trees, and SVM. The average AUC is used to identify the best model. This study utilized ADAS warnings data and driving data from the bus in Taipei metropolitan area to empirically test the aforementioned analytical methods. The results are as follows: Drivers are divided into relatively low-risk and relatively high-risk groups, and the number of ADAS warnings differs significantly between the two groups. Therefore, there is a significant difference in accident risk between the two groups. The results of the three variable selection methods are not entirely consistent. It is advisable to consider all three methods comprehensively. The final model is logistic regression due to its highest average AUC.

Keywords: accident risk, driving behavior, variable selection, statistical prediction, machine learning

目錄

摘要	i
Abstract	ii
目錄	iii
圖目錄	v
表目錄	vi
第一章 緒論	1
1.1 研究動機	1
1.2 研究目的	3
1.3 研究範圍	4
1.4 研究流程	5
第二章 文獻回顧	7
2.1 事故風險衡量指標與事故風險定義	7
2.2 先進駕駛輔助系統(ADAS)	8
2.3 事故風險影響因子	9
2.4 事故風險預測	10
2.5 小結	11
第三章 研究方法	13
3.1 資料蒐集與資料清洗	14
3.1.1 資料蒐集	14
3.1.2 國道資料刪除	14
3.1.3 空值、異常值刪除	15
3.2 事故風險衡量指標確立	15
3.2.1 事故風險群劃分	15
3.2.2 事故風險群命名	15
3.3 事故風險影響因子取用及挑選	17
3.3.1 駕駛行為特徵取用	17
3.3.2 環境因子取用	18
3.3.3 變數挑選	18
3.4 事故風險預測模型建立及挑選	18
3.5 小結	19
第四章 實例測試	20
4.1 資料蒐集與資料清洗	20
4.1.1 資料蒐集	20

4.1.2	國道資料刪除	21
4.1.3	空值、異常值刪除	22
4.2	事故風險衡量指標確立	23
4.2.1	事故風險群劃分	23
4.2.2	事故風險群命名	26
4.3	事故風險影響因子取用及挑選	27
4.3.1	駕駛行為特徵取用	27
4.3.2	環境因子取用	29
4.3.3	變數挑選	30
4.4	事故風險預測模型建立及挑選	34
4.4.1	預測模型驗證	34
4.4.2	最終模型	36
第五章	結論與建議	39
5.1	結論	39
5.2	建議	40
參考文獻		41



圖目錄

圖 1.4-1	研究流程圖	6
圖 3-1	研究方法流程圖	13
圖 4.1-1	原始資料地理資訊	21
圖 4.1-2	清洗後資料地理資訊	22
圖 4.2-1	K-means silhouette 圖	23
圖 4.2-2	分成四群時標準化警示次數盒鬚圖	24
圖 4.2-3	分成三群時標準化警示次數盒鬚圖	24
圖 4.2-4	分成兩群時標準化警示次數盒鬚圖	25
圖 4.2-5	分成兩群後各警示次數直方圖	26
圖 4.3-1	駕駛行為特徵盒鬚圖一	28
圖 4.3-2	駕駛行為特徵盒鬚圖二	28
圖 4.3-3	本研究環境因子盒鬚圖	29
圖 4.3-4	GPS 經緯度誤差示意圖	29
圖 4.3-5	疑似相對不重要之變數	30
圖 4.3-6	各複雜度下的模型準確度	32
圖 4.4-1	決策樹 5-fold 驗證之 ROC 與 AUC	34
圖 4.4-2	SVM 5-fold 驗證之 ROC 與 AUC	34
圖 4.4-3	羅吉斯迴歸 5-fold 驗證之 ROC 與 AUC	35
圖 4.4-4	判別分析 5-fold 驗證之 ROC 與 AUC	35

表目錄

表 1.1-1	111 年傷亡交通事故之前十大肇事原因	2
表 1.1-2	近十年傷亡交通事故之肇事件數與肇事原因	2
表 1.3-1	每十萬輛機動車輛事故率(件/十萬輛)	4
表 2.2-1	ADAS 功能	9
表 3.3-1	駕駛行為特徵	17
表 4.1-1	ADAS 資料欄位	20
表 4.1-2	車速車距資料欄位	21
表 4.1-3	原始車速車距敘述性統計值	22
表 4.1-4	清洗後車速車距敘述性統計值	22
表 4.2-1	兩風險群大小	26
表 4.2-2	兩風險群 t 檢定表	27
表 4.3-1	本研究取用之環境因子	29
表 4.3-2	變數重要性結果	31
表 4.3-3	變數挑選結果	33
表 4.4-1	各模型平均 AUC	36
表 4.4-2	羅吉斯迴歸結果一	36
表 4.4-3	羅吉斯迴歸結果二	37
表 4.4-4	羅吉斯迴歸結果三	37
表 4.4-5	羅吉斯迴歸結果四	38

第一章 緒論

1.1 研究動機

交通事故往往造成社會與事故當事人於多方面的損失。交通事故涵蓋之成本項目(周榮昌等, 2019)可分為人力成本, 如生產力損失成本、生活品質降低成本、醫療成本; 財物損失成本, 如車輛損失成本; 外部成本, 如延滯成本、能耗成本、汙染成本。量化上述成本, 該年我國交通事故成本高達 427,334 百萬, 佔我國 GDP 總額之 3.139%。而為了降低此衝擊, 減少交通事故為重要的課題。

事故風險預測有助於減少交通事故。事故風險預測能使主管機關了解高風險駕駛為何, 進而能對之加強教育訓練以及風險監控, 以減少交通事故。同時, 藉由預測模型的建立, 能了解高風險因子為何, 進而能避免暴露在高風險因子之下, 以減少交通事故。

進行事故風險預測時, 首先須測量出事故風險真值, 以作為後續預測模型之預測目標, 而事故數、事故率為最常見的風險衡量指標(Chen et al., 2017; Tanishita & Van, 2017; Wang et al., 2018)。不過由於交通事故並非常態性的事件, 導致建立指標所需時間漫長, 動輒數年。

除了事故數、事故率, ADAS(Advanced Driving Assistance System)警示事件同樣可用於建立風險衡量指標, 且指標建立所耗時間較短。ADAS 警示反映了不當駕駛行為的發生, 如變換車道不當、車距過短, 而這些不當駕駛行為便是常見之事故原因, 如表 1.1-1 所示, 故 ADAS 警示次數越多, 事故風險越高。相比於事故, ADAS 警示的優勢為, 發生的頻率更高, 數週內便能觀測到大量的警示, 故指標建立所耗時間較短。另因 ADAS 日漸普及, 警示資料取得已越來越容易。

「從 ADAS 警示次數直接測量事故風險」以及「事故風險預測」, 皆能找出高風險駕駛為何, 但後者所需時間更短, 且能了解高風險因子為何, 因此具有不可取代之地位。事故風險預測須訓練預測模型, 由預測模型訓練結果, 如參數, 表達事故風險影響因子與事故風險的關係, 便能了解高風險因子為何, 這是直接測量事故風險所無法達成。至於所需時間, 以 ADAS 警示次數直接測量事故風險, 其資料蒐集仍須數週; 然而事故風險預測時, 只須數小時蒐集事故風險影響因子資料(如駕駛行為、環境因子), 接著將資料輸入已訓練的預測模型, 即可預測出事故風險, 因此事故風險預測較有效率。

駕駛行為能有效地預測事故風險，因為交通事故主要由不當的駕駛行為所導致，表示駕駛行為是交通事故的重要影響因子。根據事故肇因資料（警政署統計查詢網事故概況統計，民 112 年），如表 1.1-1 所示，該年前十大肇事原因中，除了「開啟車門不當」，其餘九個皆為駕駛行為不當。將觀察期間延長，如表 1.1-2 所示，近十年傷亡交通事故同樣以駕駛人過失為主要之肇事原因，且年年都占九成以上。

表 1.1-1 111 年傷亡交通事故之前十大肇事原因

	百分比
未依規定讓車	20.03
轉彎或轉向不當（含變換車道）	15.49
未保持行車安全距離或間隔	8.66
違反號誌標誌管制	7.44
起步未注意其他車或人	3.56
未依規定減速或超速失控	1.91
逆向行駛	1.08
酒後駕車失控	1.01
違規超车或爭道行駛	0.99
開啟車門不當	0.69

資料來源：內政部警政署統計交通組

表 1.1-2 近十年傷亡交通事故之肇事件數與肇事原因

	肇事件數	駕駛人過失	行人或乘客過失	機件故障	交通管制設施缺陷	其他
101 年	249,465	244,306	3,119	856	350	834
102 年	278,388	272,541	3,635	930	348	934
103 年	307,842	301,685	3,843	985	309	1,020
104 年	305,413	299,003	3,895	946	305	1,264
105 年	305,556	299,357	3,572	936	350	1,341
106 年	296,826	291,073	3,401	797	357	1,198
107 年	320,315	313,974	3,625	922	376	1,418
108 年	341,972	335,174	4,145	922	362	1,369
109 年	362,393	355,045	4,332	1,017	401	1,598
110 年	358,221	351,600	3,852	992	390	1,387

資料來源：內政部警政署統計查詢網

事故風險的影響因子眾多，然過往研究較少重視變數選擇，但變數過多可能帶來負面影響。單從上段提及的駕駛行為方面，就能取出數十個變數。而除了駕駛行為，駕駛環境也會影響事故風險，如車流量、道路幾何、時段、天氣。變數過多時，容易導致參數標準誤大，進而導致多個變數不顯著，而不知從何刪起。因此變數選擇有其重要性。

1.2 研究目的

本研究欲提出事故風險預測之方法，方法可分為三部分，內容如下：

1. 事故風險衡量指標快速確立：本研究欲基於 ADAS 警示次數，建立事故風險衡量指標，以取代資料蒐集時間長的事故數、事故率。
2. 事故風險影響因子取用及挑選：本研究欲從駕駛行為以及行駛環境中取用若干個風險影響因子，接著以變數選擇方法，挑選出與事故風險較有關聯的因子，以提升參數有效性，並避免模型結果出現多個不顯著變數。
3. 事故風險預測模型建立及挑選：本研究欲建立多個預測模型，再利用模型品質評估指標挑選之，以找尋最適合該資料集的模型。

本研究之研究目的著重於分析方法的提出。本研究實例測試章節之 ADAS 警示事件資料的單位為每天每車，屬一天行車資料的統整結果，無法精細分別出不同時段的駕駛行為。然而一天中不同時段的行駛環境，如車流量、光線，並不相同，若無考量這些干擾因素，分析結果只能大致參考，故著重於分析方法的提出。

1.3 研究範圍

本研究所謂事故風險，只考慮發生頻率此風險構面，而不考慮嚴重性。事故嚴重性多由當事人傷亡程度判斷。國內關於事故嚴重性的判定基準仍為事故後的死亡時間，而非由醫療專業人士的評估身體各部位受傷程度後判斷，判定結果較不準確。不準確的事故嚴重性將影響研究的可靠性，故本研究暫不考慮事故嚴重性此風險構面。

本研究以一般道路上的駕駛行為作為研究範圍，而不考慮國道上的駕駛行為。因為駕駛行為影響事故風險的方式，可能因道路等級的不同而不同。比如 Wang et al.(2018) 提出，速度平均數（駕駛行為特徵之一）於高速公路及郊區道路上的影響，較於市區道路大。若鎖定單一種道路等級，便能避免此問題。又因本研究於一般道路上的資料量較大，故鎖定一般道路為研究範圍。

本研究將以大客車駕駛作為研究車種，因為大客車所乘載的乘客數較多、事故率也較高。乘客越多，發生交通事故時影響到的人員越多，甚至傷亡越多。事故率高則是根據警政署統計資料(警政署統計查詢網，民國 111 年)，以十萬輛機動車輛作為曝光量，顯示民國 101 年至 110 年大客車年年皆為事故率最高的車種，且為事故率次高之大貨車的近兩倍，故本研究將以大客車駕駛作為研究範圍。

表 1.3-1 每十萬輛機動車輛事故率(件/十萬輛)

	大客車	大貨車	小客車	小貨車	機車
101 年	4,109.59	2,178.52	1,411.30	1,993.44	836.94
102 年	4,186.48	2,313.69	1,503.64	2,150.78	1,013.15
103 年	4,840.86	2,441.17	1,600.06	2,236.77	1,175.19
104 年	4,579.52	2,222.76	1,558.93	2,129.82	1,171.76
105 年	4,749.36	2,061.18	1,592.23	2,126.99	1,152.13
106 年	4,323.15	1,986.38	1,485.27	2,053.42	1,131.68
107 年	4,230.01	2,146.84	1,544.57	2,083.95	1,238.70
108 年	4,917.69	2,245.90	1,603.86	2,150.34	1,326.08
109 年	5,141.29	2,444.61	1,652.74	2,148.21	1,415.17
110 年	4,336.80	2,541.66	1,562.90	2,145.81	1,402.34

資料來源：交通部統計查詢網

1.4 研究流程

1. 確立題目：從研究動機與研究目的出發，在釐清發想出此研究議題的起源以及預期做完研究欲達成的貢獻後，構想出大致的研究架構。
2. 文獻回顧：參考事故風險衡量指標、ADAS、事故風險影響因子以及事故風險預測的相關研究，幫助本研究測量事故風險以及預測事故風險。
3. 資料蒐集：蒐集大客車的 ADAS 警示資料以及駕駛行為資料。
4. 資料清洗：將國道資料（非研究範圍）、空值、異常值予以刪除。
5. 事故風險衡量指標確立：輸入 ADAS 警示資料至 K-means 模型，以給定各駕駛事故風險等級，作為風險衡量指標。
6. 事故風險影響因子取用及挑選：取出車距的敘述統計值(最小值、Q1、Q2、Q3、最大值、平均值、標準差、偏態、峰態)、車速的敘述統計值(最小值、Q1、Q2、Q3、最大值、平均值、標準差、偏態、峰態)，以及環境因子(上午尖峰比例、下午尖峰比例、夜晚比例等等)，接著以 t 檢定、RFE、Lasso 進行變數選擇。
7. 事故風險預測模型建立及挑選：以上述第四步驟求得之風險等級作為預測目標，並以第五步驟取用之影響因子作為預測變數，輸入羅吉斯迴歸(Logistic Regression)、決策樹(Decision Tree)、判別分析(Linear Discrimination)、支援向量機(SVM)等監督式學習模型，進行模型的訓練。建模完成後，以 AUC 作為評估模型品質的指標。
8. 結論與建議：統整提出之分析方法及實例測試之結果，並建議未來可精進的方向。

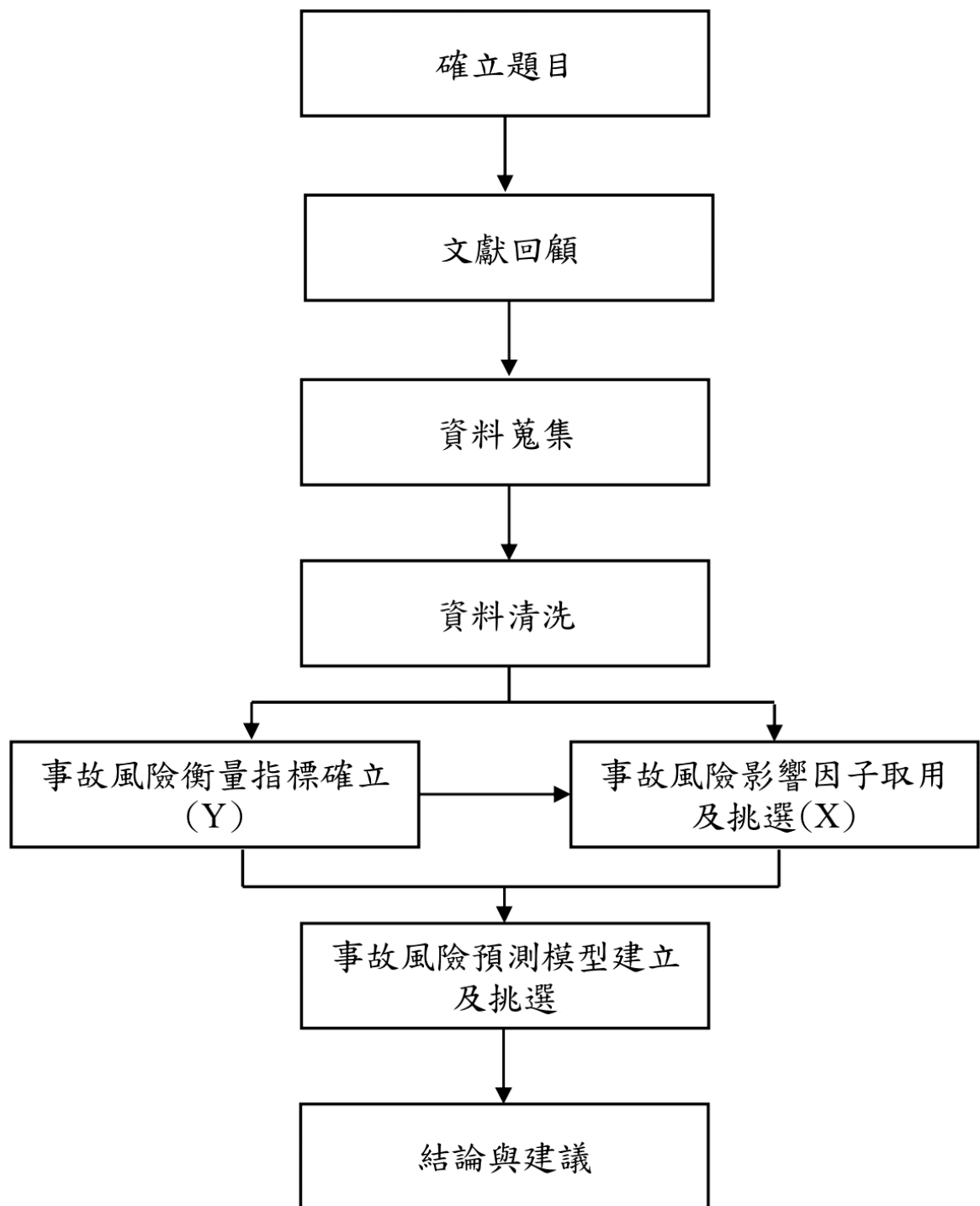


圖 1.4-1 研究流程圖

第二章 文獻回顧

本章 2.1 節回顧風險衡量指標之建構，以及該指標對事故風險的定義。建立風險衡量指標的目的為測量事故風險，最常見且最直接的風險衡量指標為事故本身，但缺點為資料蒐集時間長，動輒數年。為了降低資料蒐集所需時間，本研究將尋找其他替代指標，於本節回顧多種替代指標。而事故風險之定義，往往隨著指標的內涵不同而不同。

本章 2.2 節延續研究動機小節所提到的 ADAS 警示事件。首先簡介 ADAS 系統，了解其大致內容。接著歸納 ADAS 的眾多功能，以清楚了解 ADAS 警示事件在 ADAS 完整系統下的角色。最後聚焦於與駕駛行為有關的 ADAS 警示事件，並說明其與事故風險的關係。

本章 2.3 節利用文獻回顧，從廣大的事故風險來源，聚焦到重要的事故風險影響因子。事故風險主要來自駕駛、車流、路與環境三個層面。細部上可分為駕駛行為、駕駛經驗、駕駛態度、駕駛社會經濟特徵、駕駛生理狀態、駕駛心理狀態、擁塞程度、道路種類、道路幾何、光線、天氣等等。由於事故風險來源廣泛，故須經文獻回顧始能有架構地了解事故風險影響因子。

本章 2.4 節則對事故風險預測研究作整體性的回顧。回顧重點包含風險衡量指標、風險影響因子、預測模型。期能對事故風險的研究流程有完整性的了解。

2.1 事故風險衡量指標與事故風險定義

Xiong et al.(2019)以譜分群 (spectral clustering) 將瀕臨追撞事件分為高風險、中風險兩群，作為衡量追撞事故風險指標。此指標將風險定義為追撞事故的相對發生頻率。此研究輸入譜分群的資料為瀕臨追撞事件的減速曲線，兩瀕臨追撞事件的減速曲線越接近，越可能被分在同一群。此謂瀕臨追撞事件，指的是加速度介於-3~-4 公尺/秒平方的情況。若加速度介於-1~-2 公尺/秒平方，因減速程度不夠劇烈，故將視為一般減速狀態，而不視為關鍵事件減速狀態。

Wang et al.(2020)建構場域強度(field intensity)針對駕駛員視野可及的各個區域，給出各自的風險值，作為衡量事故風險的指標。此指標將風險定義為事故的相對發生頻率。考量的風險影響因子可分為兩大類，第一類為內部因子，包含駕駛員風險因子、道路環境風險因子、車輛風險因子。第二類為外部因子，包含靜止物品風險因子、標誌標線標誌風險因子、車輛互動風險因子、行人互動風險因子。

Karimi et al.(2021)利用 TTC(Time to Collision)作為衡量會車事故風險的指標。此指標將風險定義為會車事故的相對發生頻率。TTC 為兩車速度不變時，兩車相撞還需要的時間。若兩車逐漸遠離或者等速前進，TTC 為無限大；只有在兩車逐漸靠近時，TTC 為正數；當兩車相撞的剎那，TTC 為零。此研究以 TTC 為 3 秒作為臨界值，若大於 3 秒表示無立即的會車事故風險，小於 3 秒則表示有會車事故的風險。

Ospina-Mateus et al. (2023) 以 PET(Post Encroachment Time)作為衡量路口事故風險的指標。此指標將風險定義為路口事故的相對發生頻率。PET 為兩車通過同一空間的時間差值，通常用於評估來自不同支道的車輛於路口的事故風險。PET 值越小，表示兩車經過同一空間的時間越接近，事故風險越高。

國內亦有研究利用 TTC 以及 PET，作為衡量路口事故風險的指標。此二指標將風險定義為路口事故的相對發生頻率。此研究利用空拍機的影像資料，應用電腦視覺技術，計算出 TTC 以及 PET。最後將衝突結果繪製成圖，並堆疊於路口空拍照片上，充分將事故風險視覺化。(溫基信、王宏生、蘇志文、蔡洛緯，2019)

Dean et al. (2023) 以 MAIS(Maximum Abbreviated Injury Scale)作為衡量事故傷亡風險的指標。此指標將風險定義為事故嚴重性，也就是傷亡程度，而非較常見的事故頻率。AIS(Abbreviated Injury Scale)為評估事故當事人受傷程度的指標，將受傷嚴重度分成一到六級，數字越大嚴重越高，一級為輕傷，六級為無法醫治。而 MAIS 則是全身各部位最大的 AIS，也就是全身最嚴重的受傷程度。AIS 由汽車醫學進步協會(Association for the Advancement of Automotive Medicine)訂定，為世界公用的交通事故檢傷準則，且由醫療人員執行檢傷，相比於警察判定的受傷嚴重性，AIS 更有醫學根據。

2.2 先進駕駛輔助系統(ADAS)

ADAS 全名為 Advanced Driving Assistance System，中文名為先進駕駛輔助系統。ADAS 運用到的科技包含通訊技術，如車聯網；偵測技術，如雷達、光達、超聲波、GPS；演算法，如機器學習、深度學習、強化學習。ADAS 藉由以上科技的應用，促使駕駛過程更加安全、便利。(Nidamanuri et al., 2022)

本研究將 ADAS 依功能的差異歸納為三大類，分別為便利視覺輔助、危害警示、駕駛操作輔助。ADAS 的便利視覺輔助功能，使得駕駛人的視野更全面、更清晰 (Xu et al., 2023)。ADAS 的危害警示功能，使得駕駛人更容易掌握目前發生的危害 (Masello et al., 2023)。ADAS 的駕駛操作輔助功能，能即時避免事故發生或者簡化駕駛任務。(Abdel-Aty et al., 2022)

ADAS 的各類功能稍有差異，但皆能利用科技，提升駕駛過程的安全性及便利性。ADAS 的便利視覺輔助，利用螢幕顯示或者玻璃投影，幫助駕駛在狹窄處行駛、停車、倒車時避免擦撞周圍物體及行駛中車輛 (Kidd & McCartt, 2016)；亦能更清楚呈現行駛速度、導航資訊、路段速限等即時資訊 (Xu et al., 2023)。ADAS 的危害警示，能在危害發生時，以燈號、畫面、聲音、震動即時警示駕駛人，督促其調整駕駛行為或者檢查車輛設備，避免危害發展成事故 (Masello et al., 2023)；亦能方便掌握車輛設備狀況以及周圍物體方位 (Kidd & McCartt, 2016)。ADAS 的駕駛操作輔助，利用油門及剎車的自動控制，調整駕駛行為或者簡化駕駛任務，不只便利還能減少人為錯誤的發生 (Abdel-Aty et al., 2022)。各類 ADAS 的詳細功能統整於下表 2.2-1。

表 2.2-1 ADAS 功能

ADAS 三大功能	詳細功能	ADAS 回饋積極度	安全性提升	便利性提升
便利視覺輔助	抬頭顯示器、倒車顯影、360 度環景、導航資訊顯示、自動轉向頭燈、速限標誌辨識等等	低，僅畫面	中	高
危害警示	前方碰撞警示、跟車過近警示、車道偏移警示、超速警示、盲點偵測、停車雷達、胎壓異常警示等等	中，畫面、聲音、震動皆有可能	高	中
駕駛操作輔助	自動緊急煞車、適應性巡航、車道置中、上坡起步輔助、車身穩定系統、自動停車、自動變換車道等等	高，直接操控車輛，可能同時發出警示	高	高

ADAS 危害警示功能中，部分警示反映了駕駛行為不當，而駕駛行為不當即為事故主因，可見這些警示越多，事故風險越高。比如前方碰撞警示反映未保持安全距離；跟車過近警示也反映未保持安全距離；車道偏移警示反映變換車道不當；超速警示反映超速行為 (Ziebinski et al., 2017)。而未保持安全距離、變換車道不當、超速皆為前十大事故發生原因，如表 1-1.1 所示 (警政署統計查詢網事故概況統計，民 112)。也就是說，警示次數越多，表示不當的駕駛行為越多，進而提高事故發生頻率。

2.3 事故風險影響因子

林豐福、張開國與葉祖宏(2003)使用問卷調查資料，針對台北地區機車駕駛，蒐集有關駕駛行為、車輛設備檢查之習慣。結果顯示，年輕族群、男性、忽略車輛設備檢查者存在較高的相對事故風險。

Wickens et al.(2012)使用 2002 年 7 月至 2005 年 6 月的人口調查資料，利用羅吉特

迴歸模型，研究性別、年齡、婚姻狀況、教育程度對積極駕駛行為(driver aggression)的影響。發現男性的積極駕駛行為略高於女性，但並無顯著差異。而高齡則降低了積極駕駛行為的發生。婚姻狀況以及教育程度則與積極駕駛行為無顯著關聯。

Aurenice and Ana (2020)研究嚴重性最高的事故類型，正面對撞。指出不當超車為導致正面對撞的主要原因。本文獻讓受試者於模擬器中操作超車，作為蒐集資料之方法。受試者中男性有 55 位，女性有 25 位；年齡則介於 25-45 歲。使用 Classification and Regression Tree (CART)作為預測模型，發現此類事故的風險影響因子由重要性高到低為，駕駛類似道路環境之經驗、年齡、駕駛年數、性別、是否曾涉及事故。

Gulino et al.(2021)以車輛撞擊前的資料來預測車輛撞擊後的駕駛受傷嚴重性。而受傷程度嚴重性屬於風險構面中的嚴重性面，故受傷程度為行車風險的一部分。此研究發現車速、年齡、撞擊點是否為駕駛側、車輛登記年與駕駛受傷程度有顯著關聯性。而撞擊點是否為非駕駛側、撞擊點是否為車頭、車輛是否為掀背車、車輛是否為皮卡車、車輛是否為旅行車、車輛是否為休旅車、車輛是否為廂型車，則對駕駛受傷程度無顯著關聯性。由此篇文獻可知，車速為重要的風險影響因子。

Wu et al.(2021)研究高速公路追撞事故的風險影響因子，包含事故種類、速度樣態、道路種類、天氣、光線、駕駛分心狀況、前車動態、感知時間。研究發現，低速時的波動、減速、先加速後減速、是否位於交流道、不看行車方向會提高事故與近事故事件的發生機率。由此篇文獻可知，不只車速的高低，車速的分布（如：標準差、偏態、峰態）也是重要的駕駛行為風險影響因子。

Yu et al.(2021)利用行車軌跡預測高速公路的高風險事件，並指出何者為重要的風險因子。研究發現在事故發生前中間車道的速度、外側車道的速度標準差、外側車道的最大加速度、外側車道的最大側向加速度、外側車道的最小車距為重要的風險影響因子。由此文獻得知，不只車速，加速度、側方加速度、車距也是重要的風險影響因子。

2.4 事故風險預測

Feng and Youjia(2013) 利用 100-Car 自然駕駛研究作為資料集，以加總事故數與瀕臨事故數作為屬性，輸入 K-means 模型進行非監督式訓練，取得下一階段風險預測模型的標籤值。而風險預測模型的屬性則利用負二項模型，從性別、年齡、人格、安全相關事件作為解釋變數，並使用事故數與瀕臨事故數的加總值作為此負二項模型的反應變數，進行重要解釋變數的選取。最後利用得到的標籤值以及挑選出的重要解釋變數進行羅吉斯迴歸模型的監督式訓練。

Zhao et al.(2021)使用上海某汽車保養廠 2019 年 1 月至 2020 年 3 月的資料，利用速

度、加減速、轉彎、保養廠消費行為作為解釋變數，預測駕駛風險。以下詳細列出本篇文獻所使用之解釋變數，車速方面有以下，車速的最小值、Q1、Q2、Q3、最大值、標準差、最佳時速的占比、0-15kph 的占比、15-30kph 的占比、30-60kph 的占比、60-90kph 的占比、90-120kph 的占比、120kph 以上的占比。加減速方面有以下，每 100 公里平均的加減速次數、每 100 公里平均的輕度加減速次數、每 100 公里平均的中度加減速次數、每 100 公里平均的重度加減速次數、輕度加減速次數比例、中度加減速次數比例、重度加減速次數比例。轉彎方面有以下，右轉次數比例、左轉次數比例、迴轉次數比例、右轉均速、左轉均速、迴轉均速、右轉速度標準差、左轉速度標準差、迴轉速度標準差、右轉次數、左轉次數、迴轉次數。保養廠消費行為方面有以下，保養選擇原廠的比例、保養平均費用、保養時間間隔、保養里程間隔、保養平均購買產品數。而此篇文獻之反應變數為類別行變數，每筆資料給定無事故、一般事故、嚴重事故其中一個標籤，作為訓練模型時的標準答案。嚴重事故的定義為安全氣囊被觸動；一般事故的定義為安全氣囊未被觸動，但有人民幣 10000 元以下的維修紀錄；其餘資料則被歸為無事故。利用以上解釋變數及標籤值，訓練完羅吉斯迴歸、神經網路、隨機森林、SVM 後，發現羅吉斯迴歸有著最佳的預測力以及穩定性，而隨機森林與 SVM 的適配結果則最不穩定。

Mehdizadeh et al.(2021)使用 496 位職業大貨車司機總計超過兩千萬英里的行車資料，利用旅次特性、駕駛基本背景、天氣、駕駛行為、危險紀錄進行安全關鍵事件的預測。預測變數詳細來說有星期、是否為假日、時段、旅次長度、已行駛時間、年齡、性別、降雨機率、降雨強度、可見度、風速、平均車速、車速標準差、七天內的安全關鍵事件次數。預測模型有決策樹、廣義線性模型、SVM、XGboost 等。發現安全關鍵事件能在三十分鐘前被預測。訓練完的模型在訓練資料的蒐集時間後的兩個月都能維持穩定的預測準度。

Zhang et al.(2022)使用 2185 位上海職業大貨車司機總計 105786 趟次的車機資料以及車內影像監視資料，從旅次特性、駕駛行為、車內異常動作來預測事故風險。預測變數詳細來說有旅次長度、高速公路行駛長度、假日時間比率、夜晚時間比率、晴天時間比率、陰天時間比率、雨天時間比率、市區道路平均車速、市區道路車速標準差、高速公路平均車速、高速公路車速標準差、百公里抽菸次數、百公里打哈欠次數、百公里使用手機次數、百公里車距過短次數。風險預測模型使用零膨脹普瓦松模型。結果顯示車內異常動作為重要的預測變數，包含此類變數的模型的準度遠比不包含的佳。其中打哈欠與抽菸為嚴重影響事故風險的風險因子。

2.5 小結

替代事故數及事故率的風險衡量指標眾多，而事故風險之定義將隨著風險衡量指標不同而改變。大多相關研究的風險衡量指標將風險定義為事故發生頻率，比如：TTC、PET。但仍有部分研究的風險衡量指標將風險定義為事故嚴重性，比如：MAIS。

以駕駛行為相關之 ADAS 警示次數為資料，能測量出事故風險。因為這些警示所反映的駕駛行為，如：未保持安全距離、變換車道不當、超速，即是常見的事故原因。因此，警示次數越多，不當駕駛行為越多，越常暴露在常見事故原因下，事故風險越高。

各文獻取用之變數並非每個都顯著，其中年齡、性別、車速特徵、車距特徵，常被證明與事故風險有顯著關聯。由於車速特徵與車距特徵能表現出駕駛行為，且為文獻中常見的事故風險影響因子，因此本研究將以車速特徵與車距特徵，作為本研究所謂駕駛行為特徵。至於模型中常存在不顯著變數，則可看出變數挑選之重要性。

羅吉斯迴歸、神經網路、隨機森林、SVM 等監督式機器學習模型皆能作為事故風險預測模型。從過往事故風險預測研究中觀察到，統計預測模型能對風險因子有較深入的討論；機器學習模型則因種類繁多，而常將多個機器學習模型放入研究並互相比較預測準度。而本研究欲同時探討風險因子的影響並追求預測準度，故同時採用統計預測模型以及機器學習模型。



第三章 研究方法

本研究之研究方法主要分成四大部分，分別為：資料蒐集與資料清洗、事故風險衡量指標確立、事故風險影響因子取用及挑選、事故風險預測模型建立及挑選。參考下方研究方法流程圖，首先進行資料蒐集與資料清洗，接著左半分支為事故風險衡量指標確立，右半分支為事故風險影響因子取用及挑選，最後事故風險預測模型建立及挑選。

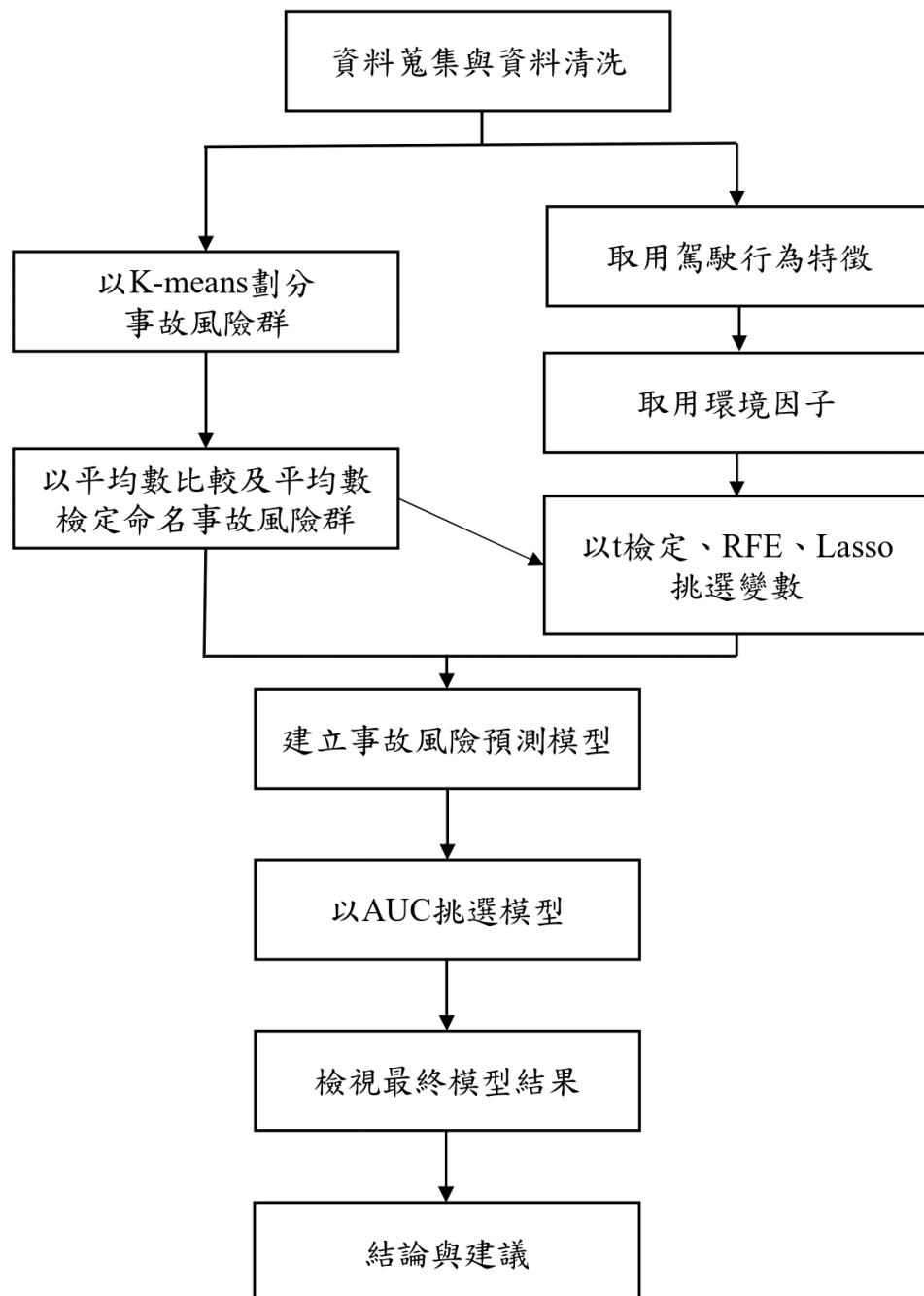


圖 3-1 研究方法流程圖

確立風險衡量指標目的為測量事故風險，產出事故風險真值，作為後續風險預測模型的預測目標(Y)。此節分為劃分事故風險群、命名事故風險群，兩個部分。劃分事故風險群時，本研究首先利用 K-means 分群法，以各類型的 ADAS 警示事件數作為輸入變數，劃分若干個事故風險群。命名事故風險群時，首先利用盒鬚圖首先大致觀察不同群之間的數值分布，接著以 t 檢定（分成兩群時）或者 ANOVA（分成兩群以上時）確認群間差異顯著。

取用及挑選風險影響因子目的為量化影響事故風險之因素，並挑選較重要之變數，作為後續風險預測模型的預測變數(X)。此節分為取用駕駛行為特徵、取用環境因子、挑選變數，三個部分。本研究取用之駕駛行為特徵有，車距以及車速分別的平均數、標準差、偏態、峰態、最小值、第一分位數、中位數、第三分位數、最大值，合計 18 個變數。接著依照資料取得程度取用環境因子，比如晨峰、昏峰、夜晚等等。最後利用 t 檢定、RFE、Lasso 三種變數挑選方法，分別排名變數之重要性，並以三方法之平均排名，作為本研究挑選變數之基準，以刪除與事故風險關聯性低的駕駛行為特徵及環境因子。

確立風險衡量指標以及取用風險影響因子後，風險預測模型所須的預測目標以及預測變數都已具備，此時便可建立風險預測模型。本研究以決策樹、向量機、羅吉斯迴歸、線性判別作為模型，以 5-fold 的平均 AUC 驗證模型品質。

3.1 資料蒐集與資料清洗

3.1.1 資料蒐集

為了以 ADAS 警示測量事故風險，須蒐集與駕駛行為相關之 ADAS 警示資料。目前具備 ADAS 警示系統之車輛，基本會有車道偏移警示、前方碰撞警示；其餘警示則因 ADAS 警示系統不同而配備有所相異，目前市面上已商用的警示還有超速警示、車距過短警示、逆向警示、闖紅燈警示、倒車車側警示等等。而警示種類越多，能捕捉到的不當駕駛行為類型越多，對事故風險的測量也越全面。

為了從駕駛行為預測事故風險，須蒐集駕駛行為資料，比如車速、車距。車距、車速能表現駕駛行為，而發生事故的頻率主要由駕駛行為決定。車速、車距亦是過往文獻常使用的駕駛行為資料，且曾被證實與事故風險有關係。

3.1.2 國道資料刪除

因應研究範圍之設定，首先刪除國道資料。但此步驟的前提是紀錄駕駛行為資料時，經緯度資料有一併紀錄。若無經緯度資料，此步驟將無法執行。

3.1.3 空值、異常值刪除

資料清洗時必須檢查各個欄位的空值，若無特別的樣態，且樣本數足夠，可直接刪除之。異常值則可分為系統預設之異常值、設備故障之異常值、違反常理之異常值。若有異常，表示此資料未能適當呈現真實情況，故予以刪除。

3.2 事故風險衡量指標確立

3.2.1 事故風險群劃分

本研究欲利用與駕駛行為相關之 ADAS 警示次數，來測量事故風險，但警示種類有很多種，該如何融合成單一的風險衡量指標？本研究以 K-means 模型，作為融合各類型警示次數的方法，方法如下：以各類型 ADAS 警示次數，作為特徵向量，輸入 K-means 模型，由模型劃分出若干群。而這若干群，將各自代表一個風險等級，比如劃分成相對高風險群、相對低風險群。此風險等級則將作為本研究之風險衡量指標。

選用 K-means 的原因為，其屬非監督式模型，訓練前不須給定標籤，切合目前無既有之風險真值的情況；且模型的訓練過程相當直覺、明確。模型訓練過程如下：先依指定群數，再隨機指派特定個群集中心，將各筆資料分至歐式距離最近的群集。接著重複更新群集中心、將資料點分在歐式距離最近的群集，直到群集中心收斂。

本研究將標準化 ADAS 警示次數，再輸入 K-means 分群，視各警示事件對事故風險的影響相同。當 K-means 使用歐式距離衡量資料點間的差異時，會受到數值尺度 (scale) 影響。舉例來說，假設超速警示較前方碰撞警示容易發生，且標準差大，若以歐式距離最小為目標，分群結果將偏重以超速警示分群，而較忽視前方碰撞警示的影響。若不標準化，等同於讓數值尺度自行決定各警示事件的權重，此方法並不合理，故應標準化警示事件次數。

K-means 於模型訓練前，須給定欲劃分的群數。而本研究將以分成一群至十群作為候選群數，以 silhouette 係數評估這十個模型的好壞，最終選擇最佳模型所對應的群數。Silhouette 係數能測量群間變異相對於群內變異有多大。Silhouette 係數數值越大，表示群間變異相較群內變異很越大，分群效果越好。Silhouette 係數詳細計算方式為如下。

$$\text{Silhouette 係數} = \frac{b-a}{\max(a,b)}$$

a: 該點與同群其他點的平均距離

b: 該點與不同群其他點的平均距離

3.2.2 事故風險群命名

完成 K-means 分群法後，各筆資料被分成若干群，此時須由研究者為每一群命名。首先觀察各群的警示次數平均數，數值較高表示該群事故風險相對高。接著檢查群間警

示次數平均數，檢定其差異有無統計上顯著的不同。若分成兩群，將使用 t 檢定檢視群間差異；若分成兩群以上，將使用 ANOVA(Analysis of Variance)檢視群間差異。若結果有顯著差異，便確立警示次數高的那群為「相對高事故風險群」，警示次數低的那群為「相對低事故風險群」。

t 檢定為比較兩個群體平均值的方法。首先以兩群變異數以及自由度建立對應之 t 分配，再求當兩群平均值無差別的機率為何。若此機率值很小（比如小於 0.05），表示若兩群平均值無差別為真，觀察到當前數值的機率很低。在統計上，觀察到當前數值的機率很低（比如小於 0.05），就能拒絕兩群平均數無差異之假設，推論兩群之平均值差異顯著。

ANOVA 為比較兩個以上群體之平均值的方法。ANOVA 的檢定值為 F 值，也就是組間變異除以組內變異的值。F 值符合 F 分布，若 F 值大得足夠極端(落於右尾 5% 機率分布面積的位置)，表示組間變異遠大於組內變異，推論此若干個群體之平均值顯著不同。



3.3 事故風險影響因子取用及挑選

3.3.1 駕駛行為特徵取用

本研究參考過往事故風險預測文獻(Wu et al., 2021)，分別對車速以及車距取最小值、Q1、Q2、Q3、最大值、平均值、標準差、偏態、峰態，車速車距各 9 個變數，共 18 個變數，欄位名稱整理於表 3.3-1。此時，車速以及車距的呈現方式不再是原始值，而是敘述統計值。

表 3.3-1 駕駛行為特徵

駕駛行為表現	駕駛行為特徵	英文名稱
車速	車速平均值	speed_mean
	車速標準差	speed_std
	車速峰態	speed_kurtosis
	車速偏態	speed_skewness
	車速最小值	speed_min
	車速第一分位數	speed_q1
	中位數	speed_q2
	車速第三分位數	speed_q3
	車速最大值	speed_max
車距	車距平均值	gap_mean
	車距標準差	gap_std
	車距峰態	gap_kurtosis
	車距偏態	gap_skewness
	車距最小值	gap_min
	車距第一分位數	gap_q1
	車距中位數	gap_q2
	車距第三分位數	gap_q3
	車距最大值	gap_max

3.3.2 環境因子取用

若不控制環境因子，恐導致模型準度下降以及參數偏誤。因為環境因子（如：道路幾何、壅塞程度、是否為夜晚等等）會影響事故風險，且與駕駛行為特徵相關。因此，雖然研究重點為駕駛行為特徵，但還須將尖峰時段比例、夜晚比例這類的環境因子放入模型，才能提升模型準度及參數不偏性。

而這裡強調取用之環境因子的數值為比例（比如：尖峰時段比例），而非類別（比如：是否為尖峰時段）。因為駕駛行為特徵為一段時間內（比如單位為每車每日）的駕駛行為統整，一段時間內的行駛環境會隨著空間及時間的改變而改變，所以並無單一的值（比如某車某日 30%時間行駛於尖峰時段，70%時間行駛於非尖峰時段），宜以比例代替類別。

3.3.3 變數挑選

變數選擇的方法(Max and Johnson, 2019)可分為過濾法(Filter)包裝法(Wrapper)、嵌入法(Embedded)三大類方法。過濾法單純考慮自變數與應變數數值之間的關係，與訓練模型無關，如利用 t 檢定了解連續型自變數與離散型應變數之間的相關性，並依此相關性排名自變數的重要程度，相關性越高越重要。包裝法利用不同自變數組合預測應變數值，能使準確度越高的自變數越重要，如 RFE(Recursive Feature Elimination)從完整的變數組合開始訓練，每次遞迴都刪掉對模型準度提升效果最差的自變數，最終以模型準度最佳的變數組合為挑選結果。嵌入法對參數最佳化時的損失函數，如平方差的和(Sum of Square)、概似(Likelihood)，加上特定的懲罰項，達到收斂參數的目的，如 Lasso 在損失函數上加上了參數絕對值和，使得各個參數都縮小，而較不重要的參數將接近零。

過濾法、包裝法、嵌入法的原理不同，變數挑選結果通常也不同，而過往相關文獻並無研究哪個方法最好。本研究將此三大類方法中各挑一個方法，分別為 t 檢定、RFE、Lasso。本研究假設三個方法權重相同，且排名間重要性差別一致，於是以平均排名作為挑選變數的參考基準，綜合考量三個方法各自產出之變數重要性。

3.4 事故風險預測模型建立及挑選

於先前階段，本研究已利用 ADAS 警示事件，求出事故風險等級，作為事故風險衡量指標；以及利用變數選擇方法，挑選出較重要的事故風險影響因子。於本階段，本研究將事故風險等級作為預測目標(Y)，將重要事故風險影響因子作為預測變數(X)，訓練事故風險預測模型。

從文獻回顧章節得知，統計預測模型以及機器學習模型各有優劣。故本研究將從統計預測模型中選用羅吉特迴歸(Logistic Regression)、判別分析(Linear Discrimination)；從

機器學習模型中選用決策樹(Decision Tree)、支援向量機(SVM)。最後以 AUC(Area Under Curve)最大的模型作為最終模型。

AUC 為衡量模型品質的指標，值為 ROC(Receiver operating characteristic curve)線之下的面積，而值越大表示模型品質越佳。ROC 為衡量模型品質的曲線，為二維座標上的線，橫坐標為真陰率 100%至 0%(等同偽陽率 0%至 100%)，縱座標為真陽率 0%至 100%。真陽率是實際為陰性的樣本數中預測成功的樣本數；真陽率是實際為陽性的樣本數中預測成功的樣本數。真陰率（橫軸）越低，真陽率（縱軸）就越高，因為預測陰性樣本的能力越差（型二誤差越大），預測陽性樣本的能力就越佳（型一誤差越小），此特性導致 ROC 為左下至右上的曲線。

3.5 小結

本研究之研究方法主要可分為資料蒐集與資料清洗、事故風險衡量指標確立、事故風險影響因子取用及挑選、事故風險預測模型建立及挑選，共四個大部分。資料蒐集與清洗包含刪除國道資料、缺失值、異常值。事故風險衡量指標確立過程包含以 K-means 劃分風險群、以平均數比較及平均數檢定為風險群命名。事故風險影響因子取用及挑選包含取用駕駛行為特徵以及環境因子，並於初步取用完畢後，續以 t 檢定、RFE、Lasso 三種變數選擇方法，決定最後取用的變數。事故風險預測模型建立及挑選包含決策樹、SVM、羅吉斯迴歸、判別分析，並以 AUC 挑選最佳模型。完成以上內容，便能快速測量出事故風險，以及從駕駛行為預測事故風險。

第四章 實例測試

4.1 資料蒐集與資料清洗

4.1.1 資料蒐集

實例測試之資料為某客運業者旗下之市區公車、公路客運、國道客運於雙北地區行駛時，產出之 ADAS 警示資料以及車距車速資料。資料中駕駛員共 441 位，車輛共 230 台，資料紀錄時間為 2022/9/24 至 2023/1/3。資料分為 ADAS 警示資料以及車速車距資料，兩份資料。

ADAS 警示資料總共 34105 筆，每天每車彙總出一筆資料，意即資料單位為日*車。ADAS 警示資料之欄位包含資料流水號、車牌號碼、資料日期、前方碰撞警示次數、車距過短警示次數、左方偏移警示次數、右方偏移警示次數、超速警示次數，如表 4.1-1。其中與駕駛行為相關之五個 ADAS 警示事件，正是本研究確立風險衡量指標的輸入資料。

表 4.1-1 ADAS 資料欄位

欄位類型	欄位名稱	欄位意義
基本背景資料	CarNo	車牌號碼
	Employee_No	駕駛員編號
	Schedule_day	資料日期
與駕駛行為相關之 ADAS 警示	FCWCount	前方碰撞警示次數
	ME_WarningCount	車距過短警示次數
	Left_LDWCount	左方偏移警示次數
	Right_LDWCount	右方偏移警示次數
	ME_OverSpeedCount	超速警示次數

上述與駕駛行為相關之 ADAS 警示事件的觸發原理如下：前方碰撞警示由過短的 TTC（相對位置除以相對速度）觸發。車距過短警示由小於 1 秒的車距（相對位置除以本車速度）觸發。左方的偏移警示由壓到或跨越左邊車道線，且未使用方向燈的行為觸發。右方的偏移警示由壓到或跨越右邊車道線，且未使用方向燈的行為觸發。超速警示則由超過標誌速限觸發。

車速車距資料總共 3,829,399 筆，感應到前方有車輛時，才會啟動系統，並記錄本車車速及車距，意即無固定之資料產出頻率。車速車距資料之每天每車紀錄資料筆數不定，筆數約為數十至數千筆，平均 251.3 筆，中位數 192 筆。車速車距資料欄位有車牌號碼、資料記錄時間（格式為年／月／日 時：分：秒）、經度、緯度、車速、車距，如表 4.1-2。其中車速與車距為主要研究對象。

表 4.1-2 車速車距資料欄位

欄位類型	欄位名稱	欄位意義
基本背景資料	CarNo	車牌號碼
	Schedule_day	資料記錄時間
車輛當下位置	Longitude	經度
	Latitude	緯度
駕駛行為	Speed	車速
	Gap	車距

4.1.2 國道資料刪除

本研究以一般道路上的駕駛行為作為研究範圍，而國道上的駕駛行為已超出研究範圍，故刪除之。因為駕駛行為特徵（如：速度平均數）與事故風險的關係，可能因道路等級不同而異。為避免此影響，本研究將以資料較多者，一般道路上的駕駛行為，作為研究範圍。共刪除 27% 資料。

下二圖為刪除國道資料前後的資料地理資訊，圖中密密麻麻的圓點為資料紀錄當下的車輛位置，紅線則為國道。考量到當今 GPS 誤差，本研究以國道周圍 150 公尺作為刪除對象。刪除後資料地理資訊如圖 4.1-2 所示。

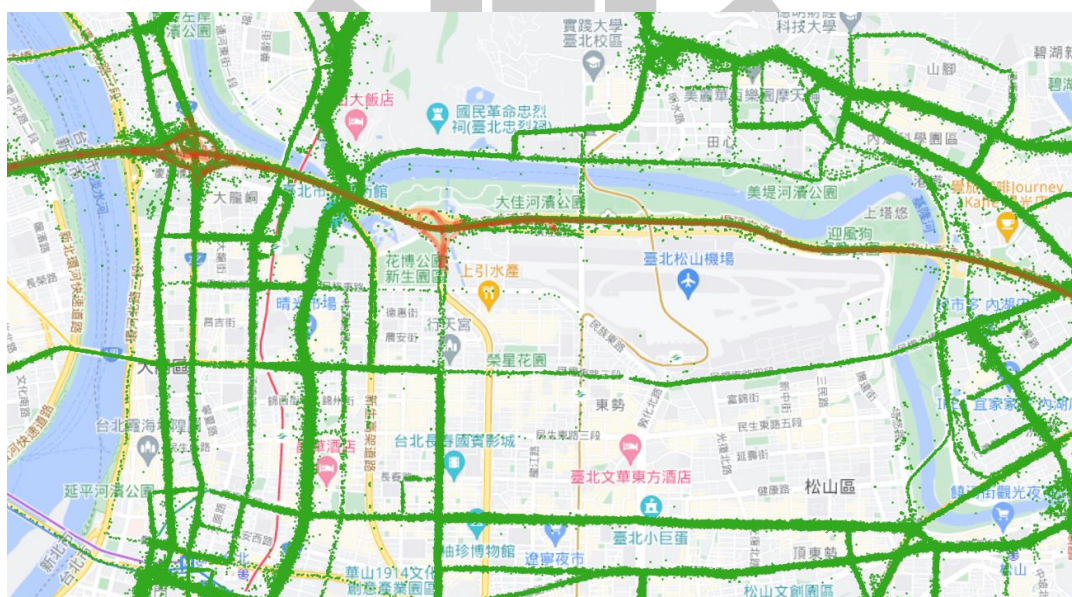


圖 4.1-1 原始資料地理資訊

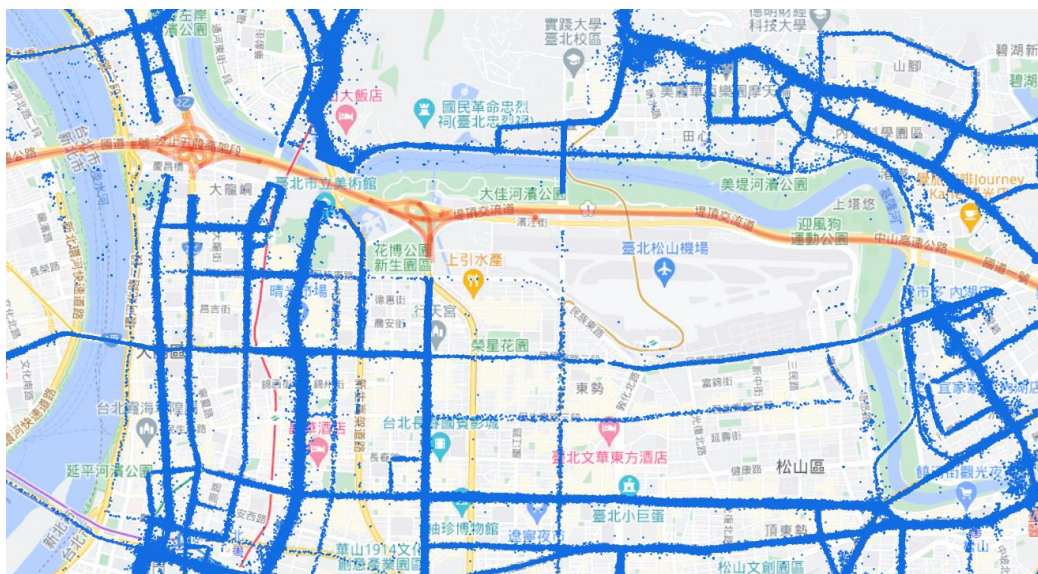


圖 4.1-2 清洗後資料地理資訊

4.1.3 空值、異常值刪除

首先處理資料中的空值。駕駛員編號之空值將予以刪除，因為若未知駕駛為何人，便無法於風險測量完成後，對該駕駛進行適合的風險管理，共刪除 0.03% 資料。而在合併 ADAS 警示資料以及車距車速資料時，因為車距車速資料比 ADAS 資料多了幾天，合併資料後會在 ADAS 的欄位留下空值，故刪除車距車速資料多出來的天數，共刪除 17% 資料。

本資料之異常值可分為：系統設定異常值、設備異常值、違反常理異常值。車距值為 1，表示系統偵測到有異常，故直接刪除，共刪除 3% 資料。某些車輛每筆車速皆為零，推估為車速紀錄設備異常，故刪除，共 0.01% 資料。本資料之違反常理異常值，包含車速過大以及車距為負，共刪除 0.03% 資料。車速過大不合理的原因為，本研究於清洗掉國道資料後，目前資料都來自一般道路，速限為 40 公里每小時至 80 公里每小時，車速超過太多並不合理，故刪除車速右尾離群值。車距合理之最小值發生於車輛碰撞時，值為 0，負值並不合理，應刪除車距負值。違反常理異常值處理前後如下表所示。

表 4.1-3 原始車速車距敘述性統計值

	最小值	Q1	中位數	平均數	Q3	最大值
車速	0	34	39	40.1	45	255
車距	-0.05	1.2	1.6	1.671	2	12.7

表 4.1-4 清洗後車速車距敘述性統計值

	最小值	Q1	中位數	平均數	Q3	最大值
車速	0	34	38	38.48	45	61
車距	0.1	1.2	1.6	1.674	2	12.7

4.2 事故風險衡量指標確立

4.2.1 事故風險群劃分

將五個 ADAS 警示次數標準化後，輸入 K-means 分群法，分成一群至十群對應之 silhouette 值如圖 4.2-1 所示。值越高，表示群間變異相對於群內變異越大，分群效果越好。由圖可知，分群效果最佳為分成四群，之後依序為分成三群、兩群、五群。分成六群以上時，分群效果明顯下降，故從候選群數中剔除。分成一群分群效果最差，因為實質上並無分群，亦從候選群數中剔除。

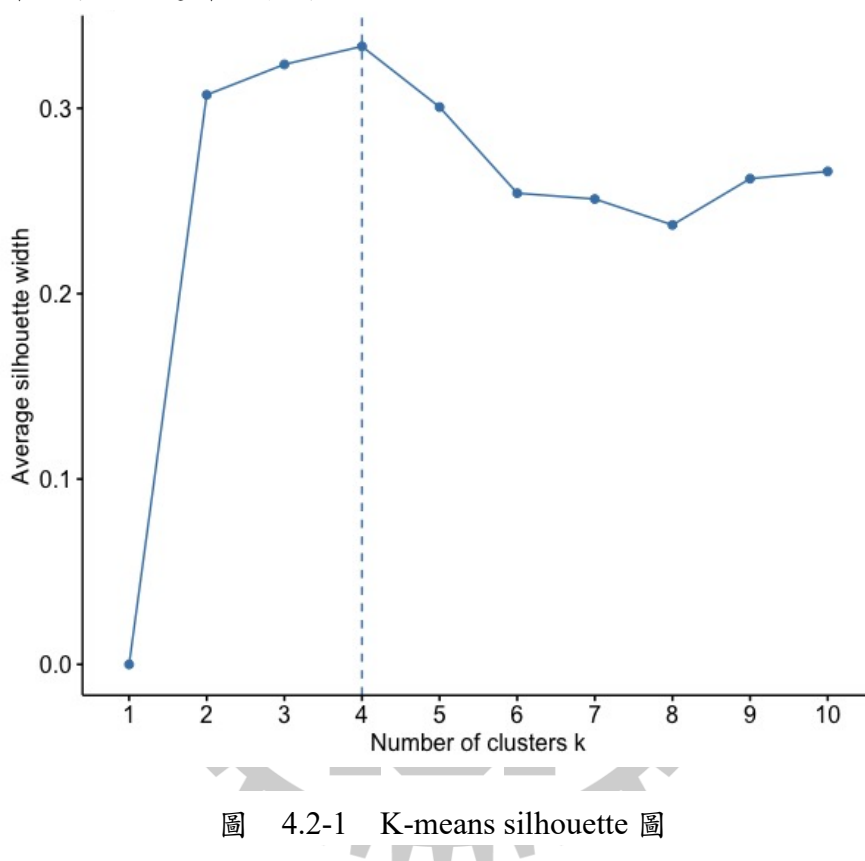


圖 4.2-1 K-means silhouette 圖

由圖 4.2-2 盒鬚圖觀察到，分成四群的結果會造成風險群命名上的困難，因為各群的警示次數沒有明顯的數值趨勢。前方碰撞警示平均次數最高者為群四；車距過近警示平均次數最高者為群一；超速警示平均次數最高者為群三，並沒有其中一群明顯為事故高風險群，也沒有明顯的特定趨勢，於是命名事故風險群時有所困難。

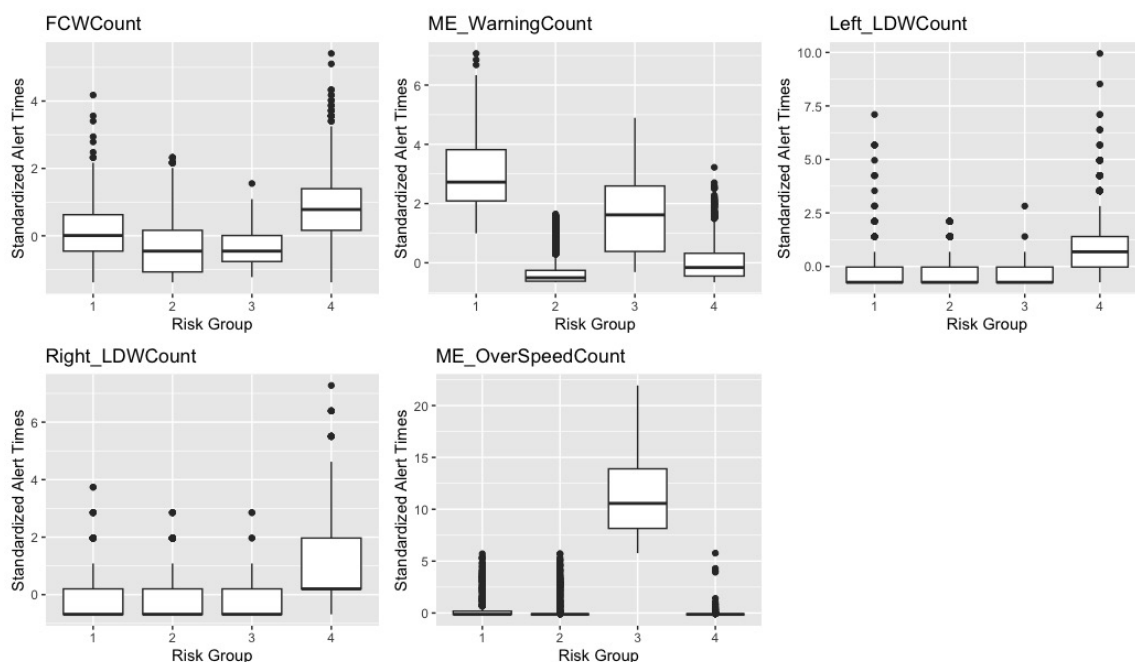


圖 4.2-2 分成四群時標準化警示次數盒鬚圖

由圖 4.2-3 盒鬚圖觀察到，分成三群時同樣存在風險群命名上的困難，因為各群的警示次數同樣沒有明顯的數值趨勢。前方碰撞警示平均次數最高者為群二；車距過近警示平均次數最高者為群一，並沒有其中一群明顯為事故高風險群，也沒有明顯的特定趨勢，於是命名事故風險群亦有所困難。

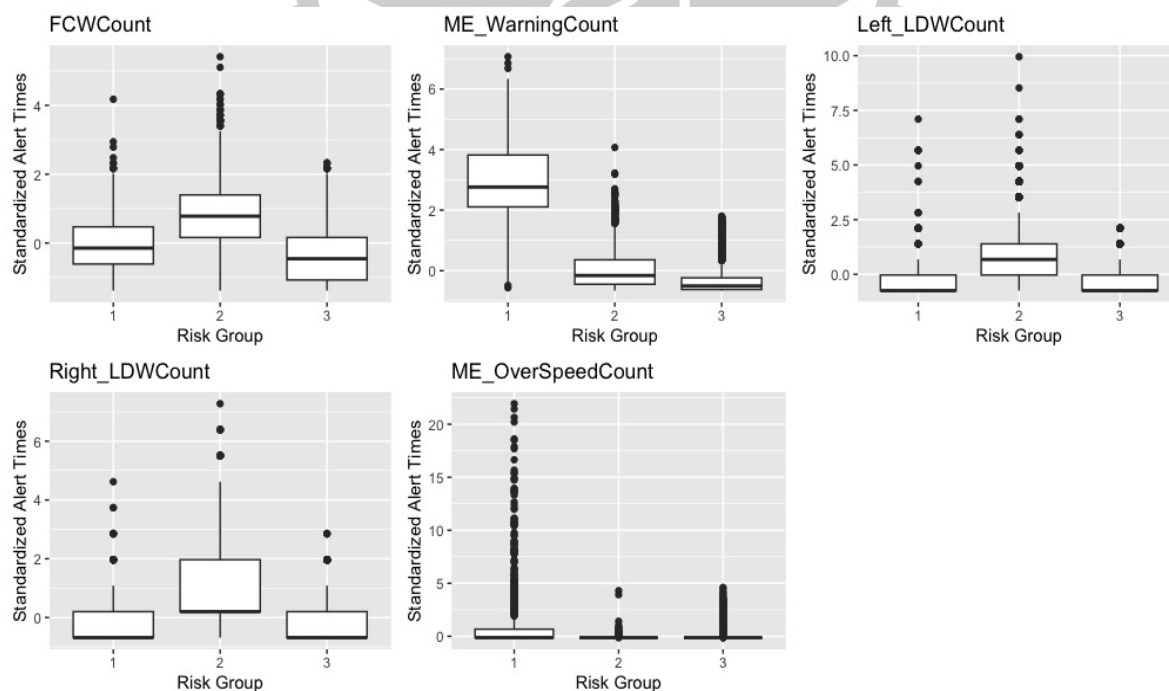


圖 4.2-3 分成三群時標準化警示次數盒鬚圖

由圖 4.2-4 盒鬚圖觀察到，分成兩群時，明顯群二的事務風險相對較高，因為各警示事件次數皆以群二最高。故本研究決定將資料分成兩個事故風險群。而群二各警示次數是否「顯著」高於群一，則於下一節由 t 檢定確認。

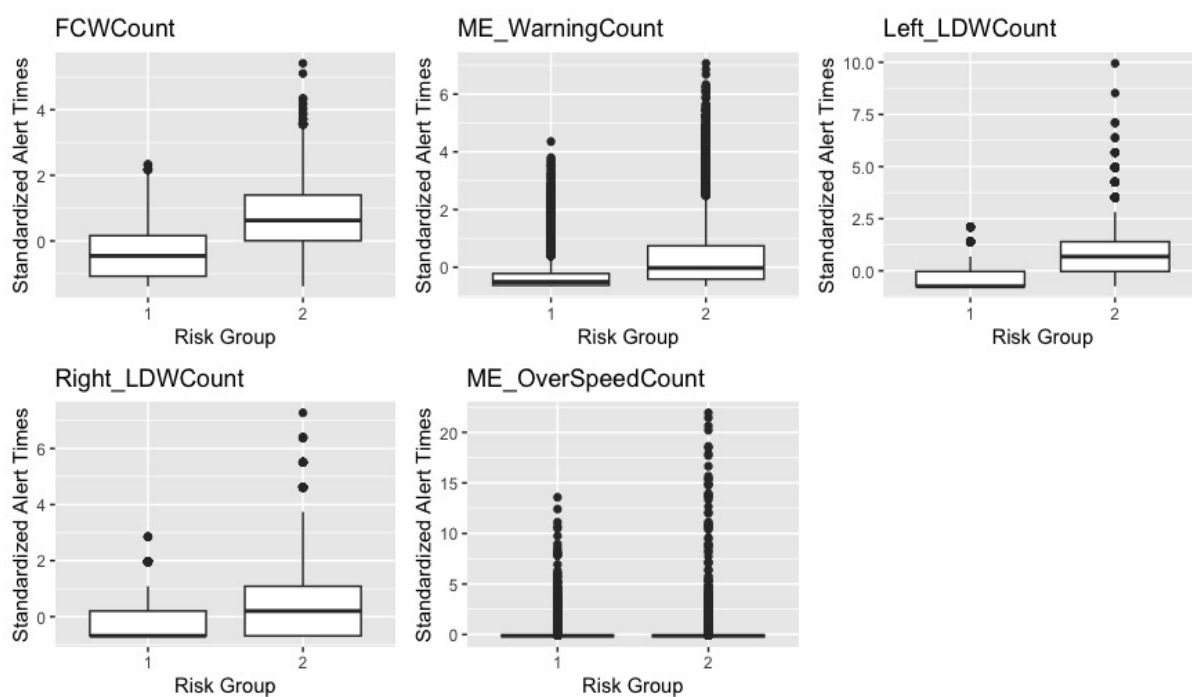


圖 4.2-4 分成兩群時標準化警示次數盒鬚圖

4.2.2 事故風險群命名

利用 K-means 完成分群並決定劃分群數後，本研究首先利用直方圖觀察警示次數未標準化時，各警示事件之分群結果。圖 4.2-5 直方圖和上一節盒鬚圖之結果皆為，群二於各警示事件之警示次數皆較群一高，故初步將群二命名為「相對高風險群」，群一則為「相對低風險群」。而從直方圖又可看出，群二之筆數較群一少，表示相對高風險群較相對低風險群少。表 4.2-1 則整理出兩群確切筆數。

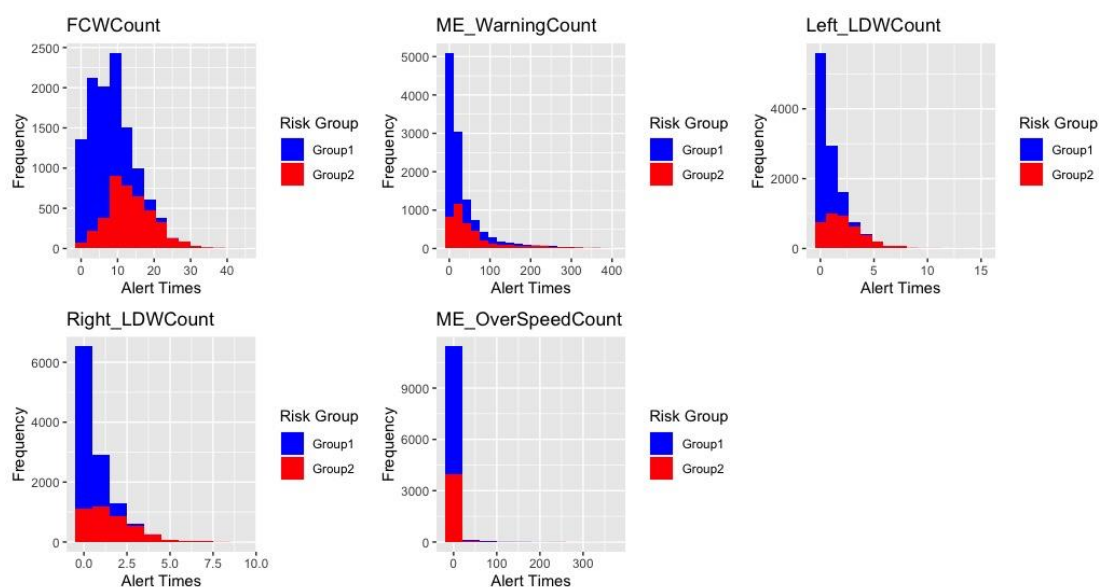


圖 4.2-5 分成兩群後各警示次數直方圖

表 4.2-1 兩風險群大小

	初步命名	筆數
群一	相對低風險群	7437
群二	相對高風險群	4258

表 4.2-2 以 t 檢定確認兩群大致具統計上之顯著差異，故最終命名群二為「相對高風險群」，群一為「相對低風險群」。前方碰撞警示次數、車距過短警示次數、左方偏移警示次數、右方偏移警示次數的 t 檢定結果顯示，p 值皆遠小於 0.05，甚至可說是接近零。p 值接近零的內涵為，若虛無假設成立（兩群的警示次數平均值相等為真），則出現當今觀察值的機率幾乎為零，故結論為拒絕虛無假設，並可推論兩群的警示次數平均值存在顯著差異。兩群之超速警示次數 t 檢定結果為，p 值等於 0.05，已相當接近差異顯著之條件（p 值小於 0.05），若樣本數再增加，p 值可能就會小於 0.05。

表 4.2-2 兩風險群 t 檢定表

ADAS 警示	相對低風險群 平均數	相對高風險群 平均數	t 值	p 值	結論
前方碰撞警示 次數	6.41	13.70	63.23	<0.001	差異顯著
車距過短警示 次數	21.17	58.83	32.65	<0.001	差異顯著
左方偏移警示 次數	0.50	2.07	55.25	<0.001	差異顯著
右方偏移警示 次數	0.37	1.54	49.86	<0.001	差異顯著
超速警示次數	1.82	2.49	1.96	0.05	差異不顯著

4.3 事故風險影響因子取用及挑選

4.3.1 駕駛行為特徵取用

為配合 ADAS 警示資料的資料單位（每天每車），駕駛行為特徵亦同樣以每天為車速車距統整單位，對每車取出當天的駕駛行為特徵，於是資料單位同為每天每車。實例測試之車速車距資料一天有數十至數千筆不等，以每天每車為單位計算車速之最小值、Q1、Q2、Q3、最大值、平均值、標準差、偏態、峰態，以及車距之最小值、Q1、Q2、Q3、最大值、平均值、標準差、偏態、峰態，共 18 個駕駛行為特徵。兩風險群分別之駕駛行為特徵分布如下圖。

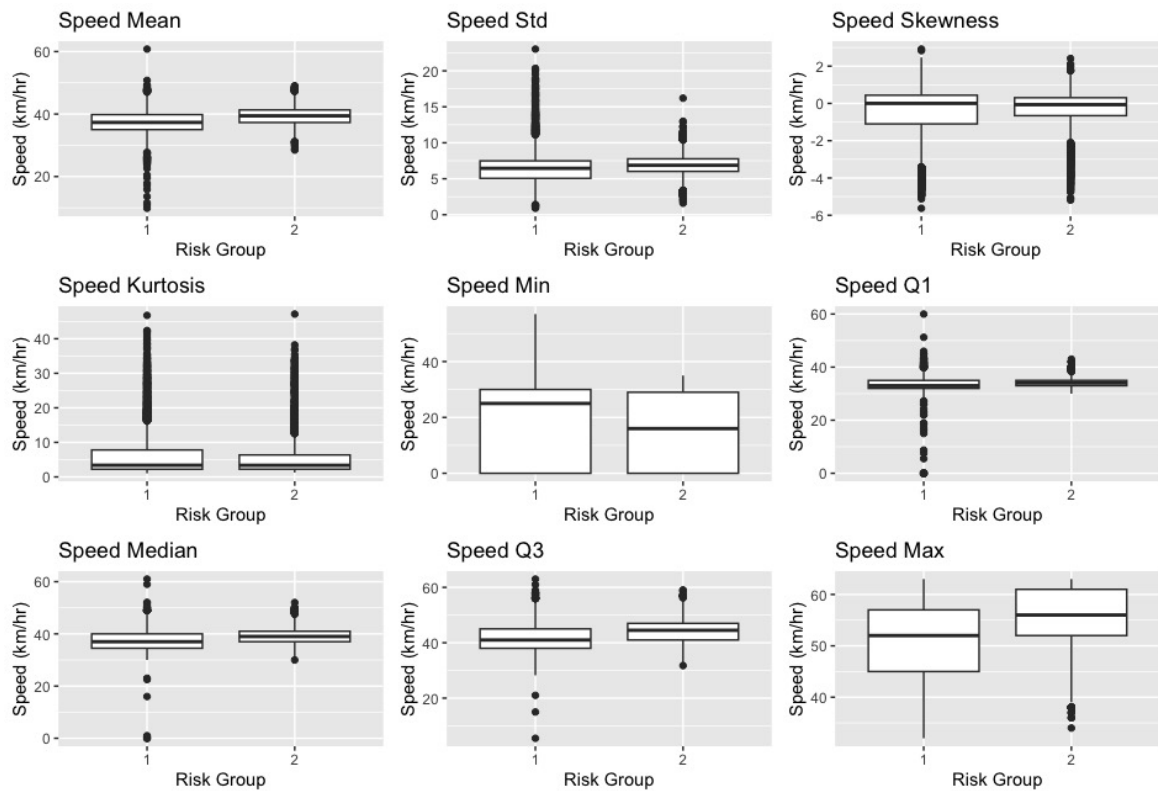


圖 4.3-1 駕駛行為特徵盒鬚圖一

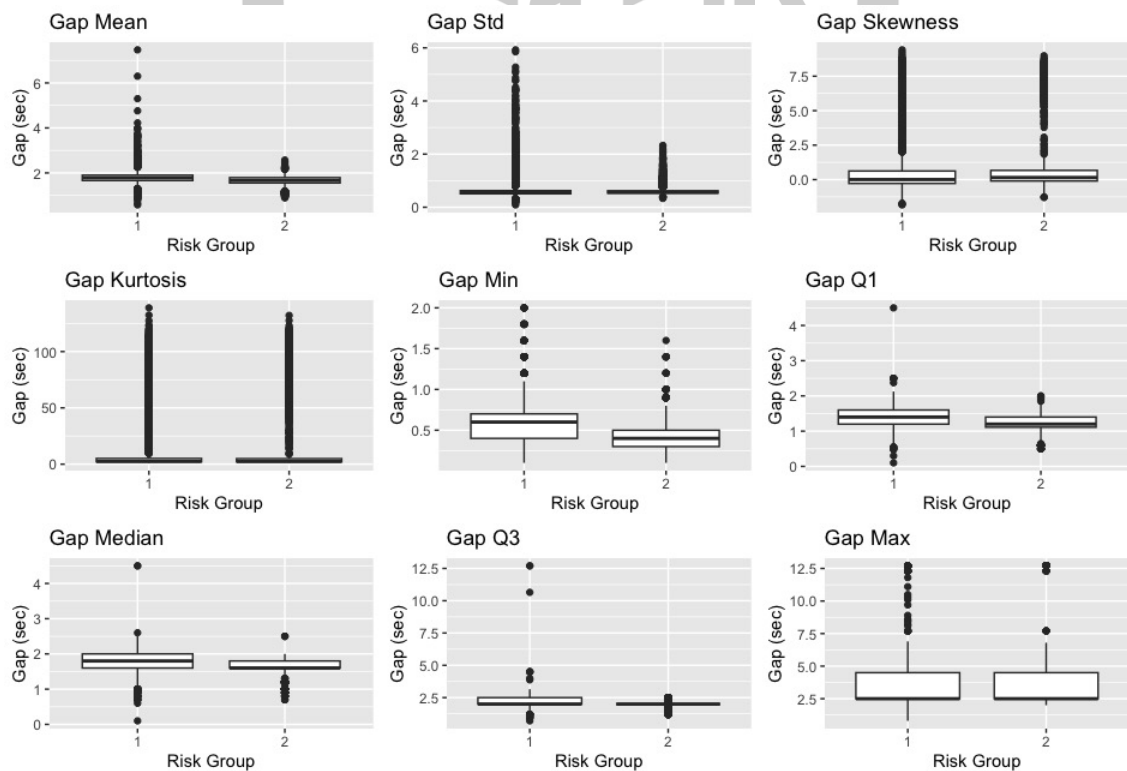


圖 4.3-2 駕駛行為特徵盒鬚圖二

4.3.2 環境因子取用

本研究取用之環境因子有上午尖峰比例、下午尖峰比例、夜晚比例，如表 4.3-3；兩風險群之環境因子數值分布則繪於圖 4.3-3。時段、天氣、道路幾何、道路等級等皆為影響事故風險的環境因子，但本研究受制於經緯度之誤差，難以準確定位車輛位置，導致無法取用道路方面的環境因子。將資料點繪製於地圖上，可以發現部分資料點出現在建築物上、河流上以及其他非道路的區域，故推估此 GPS 系統有明顯誤差，如圖 4.3-4 所示，部分經緯度位於河流、公園內。在台北市道路密集且公車路線眾多的背景下，本研究從此 GPS 系統蒐集之經緯度，無法清楚定位出車輛行駛於哪條道路，故無法取用道路方面之環境因子。

表 4.3-1 本研究取用之環境因子

資料種類	欄位名稱	欄位意義
環境因子	Morning_peak	(上班日) 上午尖峰比例
	Evening_peak	(上班日) 下午尖峰比例
	Night	夜晚比例

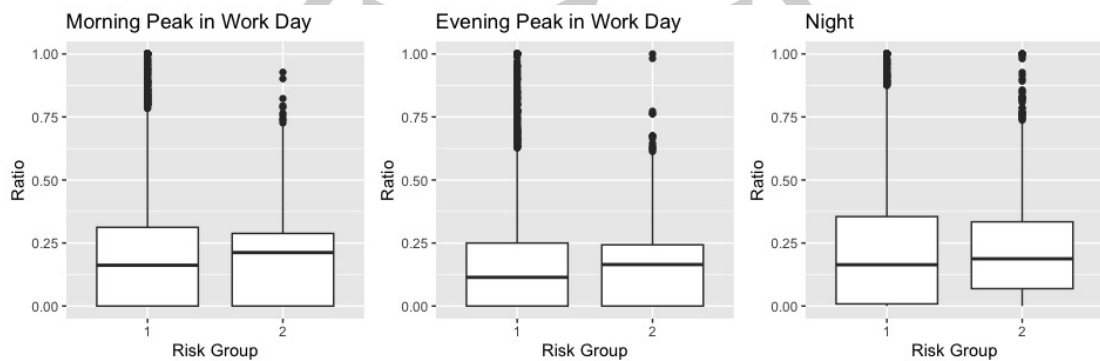


圖 4.3-3 本研究環境因子盒鬚圖



圖 4.3-4 GPS 經緯度誤差示意圖

4.3.3 變數挑選

從前兩節各變數之盒鬚圖，發現部分變數兩風險群分布相近，不因風險群不同而數值分布不同，此現象可能導致該變數預測能力有限。前兩節以盒鬚圖，查看 18 個駕駛行為特徵與 3 個環境因子，於兩風險群之數值分布有何不同。結果發現車速偏態、車速峰態、車距平均、車距峰態、上午尖峰比例、下午尖峰比例的兩風險群盒鬚圖沒有明顯不同，如圖 4.3-5 所示。

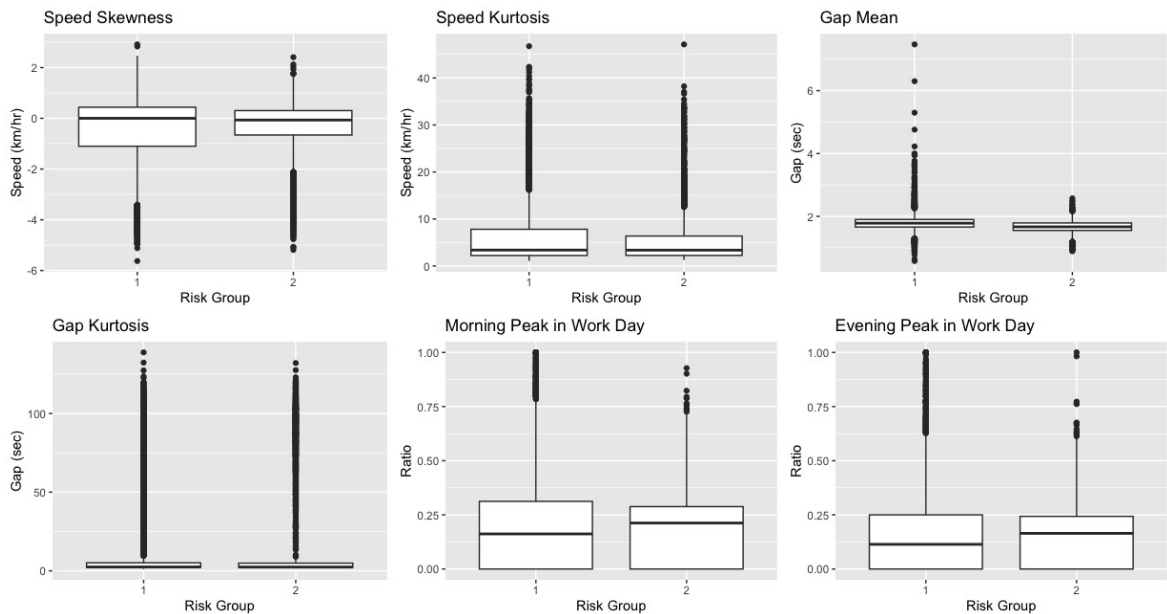


圖 4.3-5 疑似相對不重要之變數

為了更精確地挑選變數，本研究利用 t 檢定、RFE、Lasso 進行變數選擇。因為盒鬚圖能只了解數值大致的分布，單憑盒鬚圖視覺判斷變數的重要性並不準確。於是本研究利用 t 檢定、RFE、Lasso 三方法，各自排序變數重要性後，以平均變數重要性排名作為變數挑選的基準。變數挑選結果如表 4.3-4。

三種變數選擇方法衡量變數重要性的指標如下。t 檢定由 p 值衡量變數重要性。p 值越小，表示該變數之兩群平均數的差異越顯著，與事故風險越有關聯，變數重要性越高。RFE 由歷經的反覆(iteration)次數衡量變數重要性。在越後期的反覆才被刪除，表示該變數在模型準度上的貢獻越大，變數重要性越高。Lasso 由係數絕對值衡量變數重要性。Lasso 演算法自動會降低相對不重要之變數的係數絕對值，故係數絕對值越大，表示變數重要性越高。

表 4.3-2 變數重要性結果

變數	p 值	t 檢定 排名	歷經 反覆數	RFE 排 名	係數 絕對值	Lasso 排名	平均 排名
gap_min	<0.001	1	94	3	1.16	3	2.33
gap_q1	<0.001	2	75	7	1.64	2	3.67
speed_std	<0.001	11	105	1	0.58	4	5.33
speed_max	<0.001	5	100	2	0.11	13	6.67
gap_std	<0.001	14	77	6	1.97	1	7.00
speed_q1	<0.001	9	91	4	0.27	9	7.33
gap_mean	<0.001	3	24	17	0.53	5	8.33
speed_mean	<0.001	4	46	13	0.13	11	9.33
gap_max	<0.001	16	85	5	0.28	8	9.67
speed_q3	<0.001	6	49	12	0.10	14	10.67
speed_skewness	<0.001	18	68	9	0.35	6	11.00
gap_q3	<0.001	10	37	15	0.24	10	11.67
night	0.20	19	50	11	0.35	7	12.33
speed_min	<0.001	12	70	8	0.02	19	13.00
gap_skewness	<0.001	15	33	16	0.12	12	14.33
speed_kurtosis	<0.001	17	53	10	0.03	18	15.00
gap_q2	<0.001	8	7	21	0.04	17	15.33
gap_kurtosis	<0.001	13	42	14	0.01	20	15.67
speed_q2	<0.001	7	13	20	0.00	21	16.00
workpeak2	0.39	20	19	18	0.05	15	17.67
workpeak1	0.52	21	17	19	0.04	16	18.67

RFE 搭配 5-fold 驗證顯示，17 個變數時的模型準度最佳，故本研究將挑選前 17 個重要的變數進入後續的預測模型。圖 4.3-6 展示不同模型複雜度下的平均模型準度，橫軸為模型複雜度，縱軸為準確度。

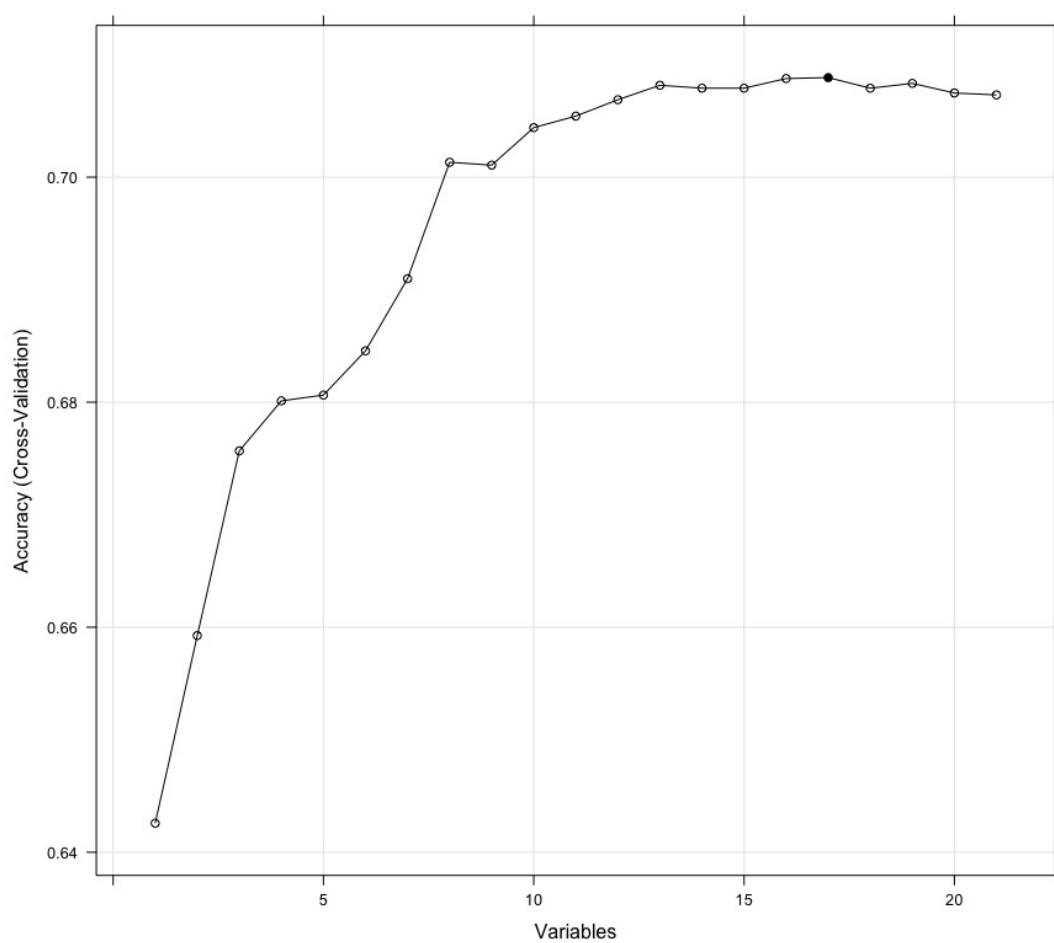


圖 4.3-6 各複雜度下的模型準確度

t 檢定、RFE、Lasso 之變數平均重要性排名如表 4.3-5，前 17 個重要的變數依序為車距最小值、車距 Q1、車速標準差、車速最大值、車距標準差、車速 Q1、車距平均值、車速平均值、車距最大值、車速 Q3、車速偏態、車距 Q3、夜晚比例、車速最小值、車距偏態、車速峰態、車距 Q3。相對不重要的四個變數車距峰態、車速 Q2、下午尖峰比例、上午尖峰比例將刪除。

表 4.3-3 變數挑選結果

變數	平均排名	重要性	是否選擇該變數
gap_min	2.33	1	是
gap_q1	3.67	2	是
speed_std	5.33	3	是
speed_max	6.67	4	是
gap_std	7.00	5	是
speed_q1	7.33	6	是
gap_mean	8.33	7	是
speed_mean	9.33	8	是
gap_max	9.67	9	是
speed_q3	10.67	10	是
speed_skewness	11.00	11	是
gap_q3	11.67	12	是
night	12.33	13	是
speed_min	13.00	14	是
gap_skewness	14.33	15	是
speed_kurtosis	15.00	16	是
gap_q2	15.33	17	是
gap_kurtosis	15.67	18	否
speed_q2	16.00	19	否
workpeak2	17.67	20	否
workpeak1	18.67	21	否

4.4 事故風險預測模型建立及挑選

4.4.1 預測模型驗證

實例測試以羅吉斯迴歸作為最終模型。從各模型 5-fold 驗證過程的五個 AUC 得知，羅吉斯迴歸的平均 AUC 最高，表示模型品質最佳。依模型品質排名，接續為判別分析、決策樹、SVM。

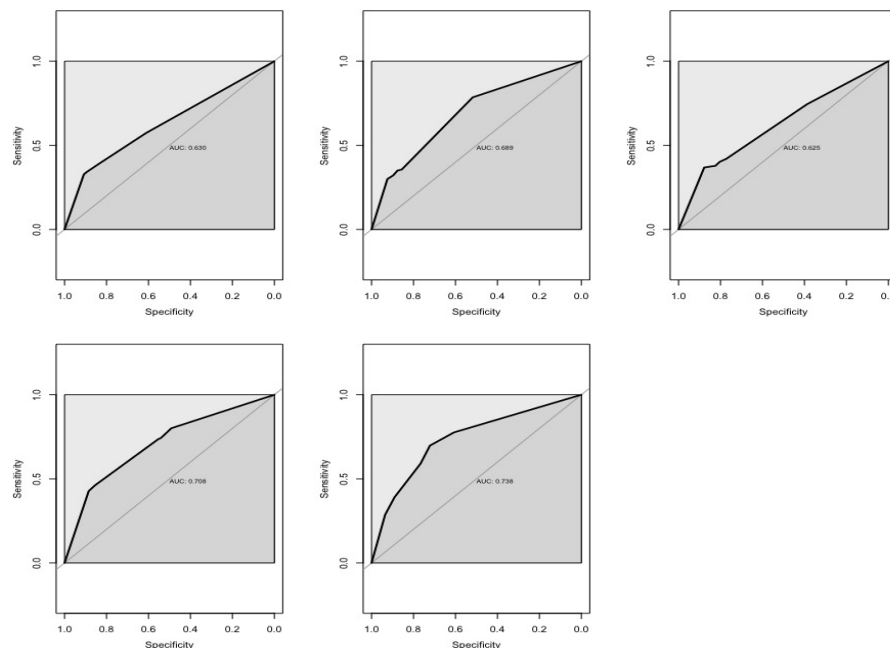


圖 4.4-1 決策樹 5-fold 驗證之 ROC 與 AUC

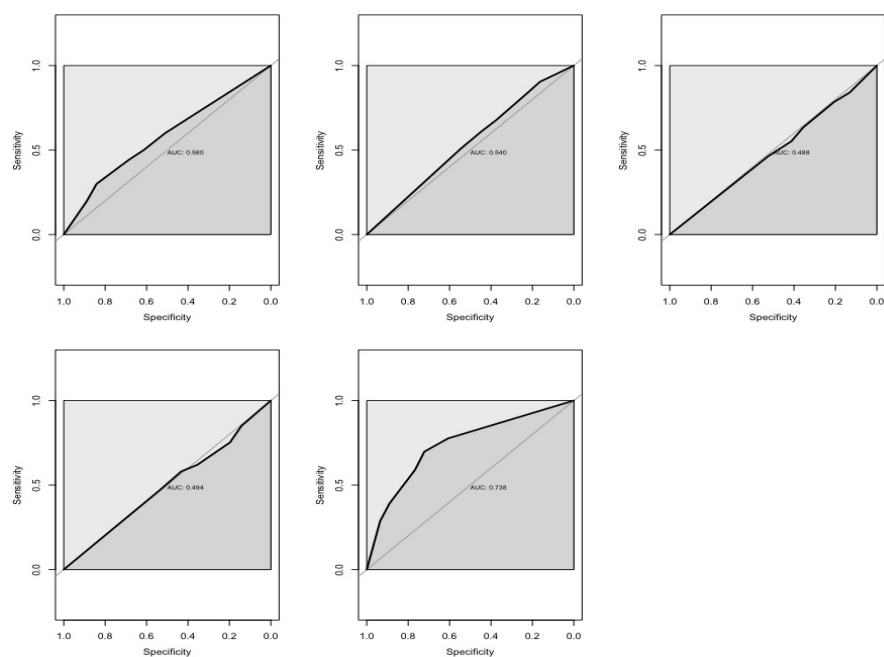


圖 4.4-2 SVM 5-fold 驗證之 ROC 與 AUC

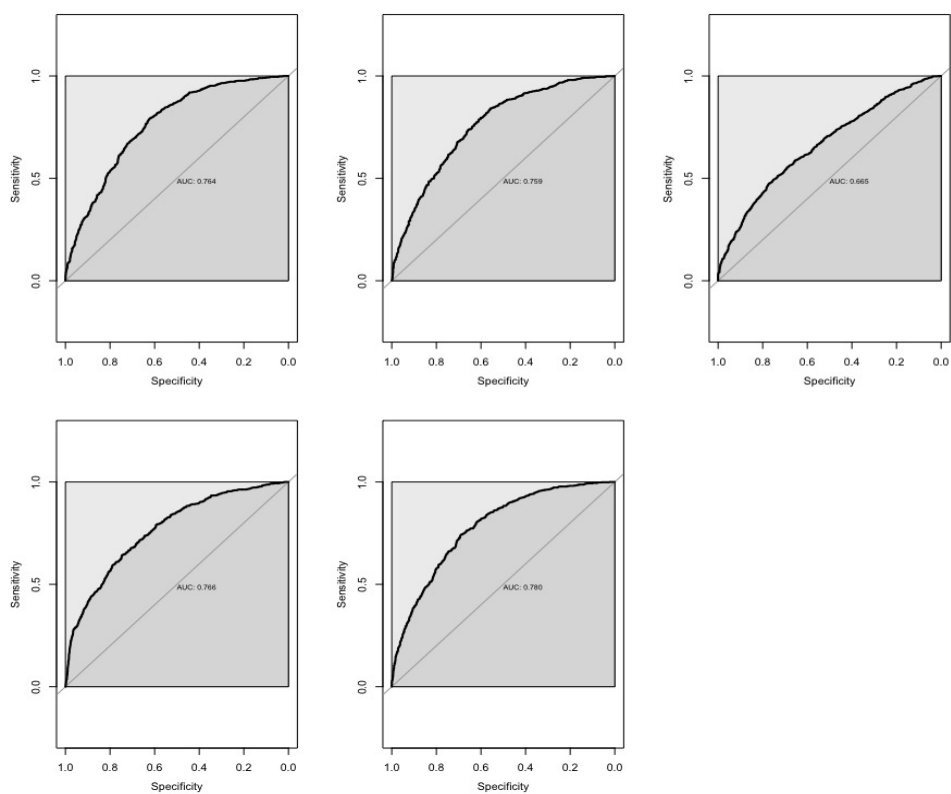


圖 4.4-3 羅吉斯迴歸 5-fold 驗證之 ROC 與 AUC

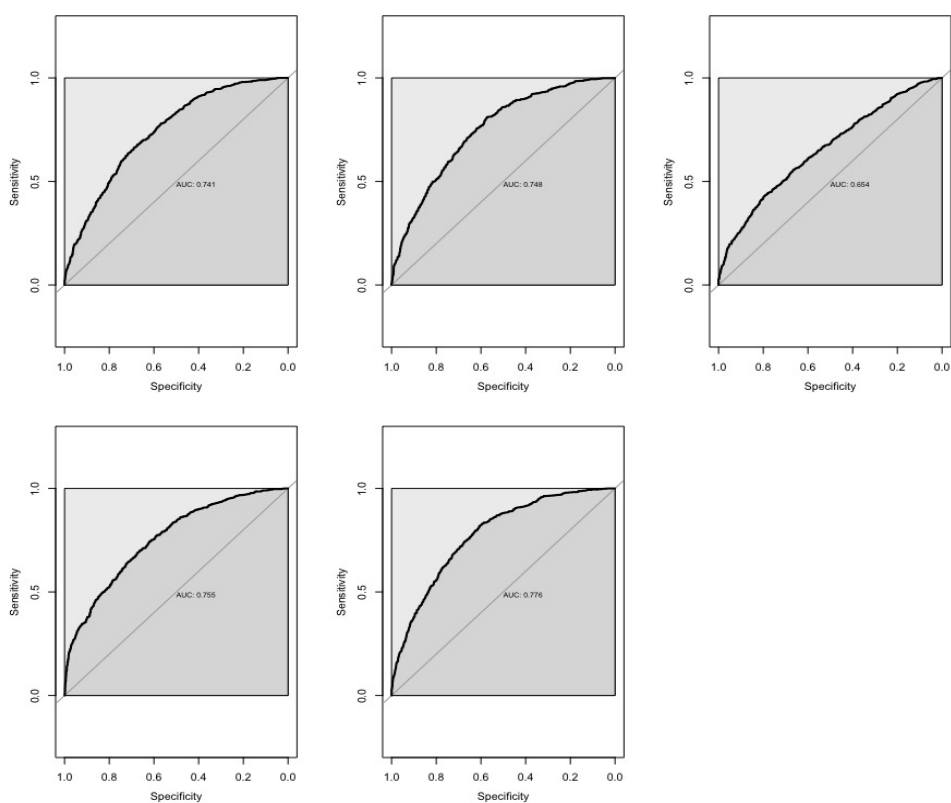


圖 4.4-4 判別分析 5-fold 驗證之 ROC 與 AUC

表 4.4-1 各模型平均 AUC

	決策樹	SVM	羅吉斯迴歸	判別分析
平均 AUC	0.678	0.568	0.747	0.735

4.4.2 最終模型

羅吉斯迴歸的模型結果如下表，挑選出的 17 個變數中（不含截距項），車距平均值、車距 Q2、車距 Q3 的 p 值大於 0.05。表示不拒絕虛無假設（該變數參數為零為真），也就是此三變數對事故風險影響不顯著。而本研究首先刪除 p 值最大之車距 Q2，再重新適配模型。

表 4.4-2 羅吉斯迴歸結果一

	參數	標準誤	z 值	p 值
(Intercept)	0.740	0.676	1.096	0.273
speed_mean	0.149	0.059	2.538	0.011
speed_std	-0.600	0.037	-16.150	0.000
speed_skewness	-0.395	0.072	-5.478	0.000
speed_kurtosis	-0.034	0.011	-2.990	0.003
speed_min	-0.015	0.003	-5.312	0.000
speed_q1	-0.282	0.034	-8.314	0.000
speed_q3	0.089	0.033	2.748	0.006
speed_max	0.118	0.009	12.438	0.000
gap_mean	-0.537	0.567	-0.946	0.344
gap_std	-1.579	0.245	-6.449	0.000
gap_skewness	-0.248	0.050	-4.928	0.000
gap_min	-1.065	0.120	-8.860	0.000
gap_q1	-1.629	0.287	-5.683	0.000
gap_q2	-0.142	0.290	-0.488	0.626
gap_q3	0.228	0.149	1.531	0.126
gap_max	0.276	0.041	6.810	0.000
night	-0.357	0.108	-3.308	0.001

本研究重複刪除不顯著變數並重新適配模型，直到變數都顯著後才停止。刪除車距 Q2 後，車距平均值以及車距 Q3 仍然不顯著，如表 4.2-2，故刪除車距 Q3，再重新適配模型。刪除車距 Q3 後，車距平均值仍然不顯著，如表 4.2-3，故刪除車距平均值，再重新適配模型。

表 4.4-3 羅吉斯迴歸結果二

	參數	標準誤	z 值	p 值
(Intercept)	0.760	0.675	1.126	0.260
speed_mean	0.149	0.059	2.542	0.011
speed_std	-0.600	0.037	-16.157	0.000
speed_skewness	-0.396	0.072	-5.481	0.000
speed_kurtosis	-0.034	0.011	-2.985	0.003
speed_min	-0.015	0.003	-5.313	0.000
speed_q1	-0.282	0.034	-8.317	0.000
speed_q3	0.089	0.033	2.752	0.006
speed_max	0.118	0.009	12.442	0.000
gap_mean	-0.712	0.439	-1.623	0.105
gap_std	-1.546	0.236	-6.561	0.000
gap_skewness	-0.247	0.050	-4.911	0.000
gap_min	-1.057	0.119	-8.876	0.000
gap_q1	-1.625	0.287	-5.671	0.000
gap_q3	0.228	0.149	1.531	0.126
gap_max	0.275	0.040	6.794	0.000
night	-0.357	0.108	-3.302	0.001

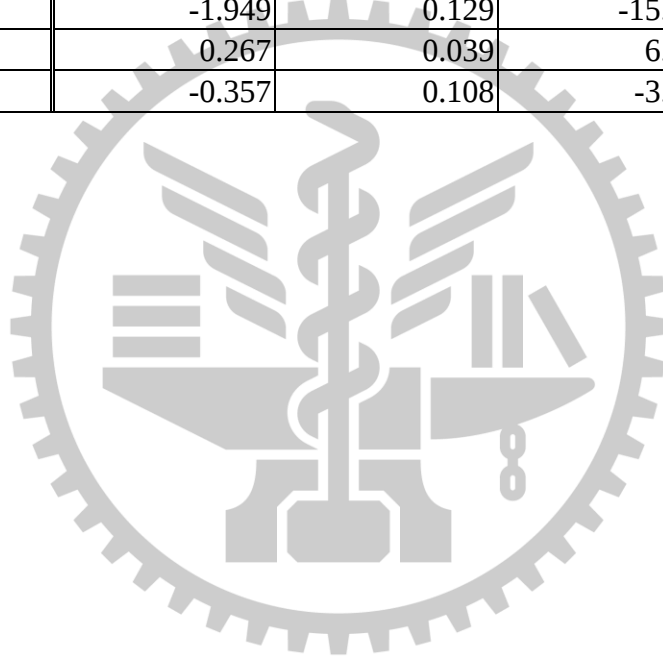
表 4.4-4 羅吉斯迴歸結果三

	參數	標準誤	z 值	p 值
(Intercept)	0.797	0.673	1.185	0.236
speed_mean	0.151	0.059	2.566	0.010
speed_std	-0.596	0.037	-16.101	0.000
speed_skewness	-0.393	0.072	-5.443	0.000
speed_kurtosis	-0.033	0.011	-2.942	0.003
speed_min	-0.015	0.003	-5.277	0.000
speed_q1	-0.282	0.034	-8.316	0.000
speed_q3	0.088	0.032	2.716	0.007
speed_max	0.117	0.009	12.385	0.000
gap_mean	-0.394	0.386	-1.020	0.308
gap_std	-1.571	0.233	-6.730	0.000
gap_skewness	-0.250	0.050	-4.978	0.000
gap_min	-1.072	0.119	-9.040	0.000
gap_q1	-1.692	0.283	-5.976	0.000
gap_max	0.276	0.040	6.826	0.000
night	-0.356	0.108	-3.294	0.001

在刪除完車距 Q2、車距 Q3 後、車距平均值，留下的 14 個變數（不含截距項）p 值皆小於 0.05，拒絕虛無假設（該變數參數為零為真），表示這 14 個變數皆對事故風險有顯著的影響。而此模型的截距項也是適配的四個模型中最不顯著的（p 值最大），表示事故風險已大多由其他變數解釋，沒解釋到的部分很少。

表 4.4-5 羅吉斯迴歸結果四

	參數	標準誤	z 值	p 值
(Intercept)	0.492	0.605	0.813	0.416
speed_mean	0.154	0.059	2.630	0.009
speed_std	-0.592	0.037	-16.093	0.000
speed_skewness	-0.387	0.072	-5.382	0.000
speed_kurtosis	-0.033	0.011	-2.886	0.004
speed_min	-0.015	0.003	-5.257	0.000
speed_q1	-0.281	0.034	-8.304	0.000
speed_q3	0.087	0.032	2.693	0.007
speed_max	0.116	0.009	12.389	0.000
gap_std	-1.663	0.216	-7.698	0.000
gap_skewness	-0.236	0.048	-4.884	0.000
gap_min	-1.098	0.116	-9.476	0.000
gap_q1	-1.949	0.129	-15.161	0.000
gap_max	0.267	0.039	6.766	0.000
night	-0.357	0.108	-3.308	0.001



第五章 結論與建議

5.1 結論

1. 從 ADAS 原始警示次數分布得知，許多觀察值非零，表示於資料蒐集的三個多月內，已觀測到許多 ADAS 警示發生。
2. 本研究以 silhouette 係數評估分群效果的結果，作為劃分群數參考；並以盒鬚圖比較各群平均數的結果，作為風險群命名參考。實例測試中，本資料集分群效果最佳為分成四群，之後依序為分成三群、兩群。但分成四群或三群時，各群各 ADAS 警示次數之平均數並無明顯趨勢；分成兩群時，群二明顯各 ADAS 警示次數高於群一，因此決定分成兩群，並初步命名群二為相對高風險群，群一為相對低風險群。
3. 風險群分成兩群時本研究以 t 檢定確認分群後各群平均數之差異具顯著性，分成兩群以上則使用 ANOVA，確定差異顯著後才確立風險群名稱。實例測試中，資料分為兩群，以 t 檢定結果，確認兩群前方碰撞警示、車距過短警示、左方偏移警示、右方偏移警示之平均次數，有顯著差異；超速警示平均次數則接近有顯著差異。因此確定命名群二為相對高風險群，群一為相對低風險群。
4. 本研究使用三種變數挑選方法，t 檢定、RFE、Lasso，挑選已取用之變數。實例測試中，三種方法的結果並不相同。比如平均排名最高的變數，車距最小值，在 t 檢定中排名第一，在 RFE、Lasso 中皆排名第三。
5. 變數挑選之後，預測模型中仍可能出現不顯著之變數。實例測試中，已從原始 21 個變數中刪除 4 個較不重要之變數，但將剩下的 17 個變數輸入羅吉斯迴歸，訓練結果顯示，仍有三個變數不顯著。
6. 本研究使用羅吉斯迴歸、判別分析、決策樹、SVM，作為預測模型。實例測試中，四模型之平均 AUC 不同，其中羅吉斯迴歸平均 AUC，表示模型品質最佳。
7. 駕駛行為有預測事故風險的能力。實例測試中，羅吉斯迴歸的訓練結果顯示，大多駕駛行為特徵與事故風險具顯著關係。

5.2 建議

1. 須注意不同表格之資料單位。本研究使用 ADAS 警示資料以及車速車距資料，前者資料單位為每車每天，後者資料單位不定。本研究於合併表格時，決定共同以每車每天作為資料單位，後續才能一同分析。
2. 資料單位可更精細。本研究之資料單位為每天每車，但一天取一筆資料恐怕過度模糊化駕駛行為。
3. 若取用環境因子，須注意該資料單位下，是否能取得環境因子唯一值，是否需要以比例代替。本研究使用之資料單位為每車每天，若想取用是否為夜晚、是否為尖峰時段這型的類別變數時，會發現沒有唯一值存在，因為一天的資料往往部分來自白天、部分來自夜晚；部分來自尖峰時段、部分來自非尖峰時段。本研究因此使用夜晚比例、尖峰時段比例來代替是否為夜晚、是否為尖峰時段。
4. 若欲取用環境因子時用到經緯度資料，比如欲加入道路環境之欄位，須注意該 GPS 系統之誤差。若誤差相對於路網密度很大，則不適合使用此經緯度辨識車輛位於哪條路，更無法得知位於哪個車道。
5. 環境因子可取用更多。過往文獻提到車流環境、道路環境皆會影響事故風險，但本研究因資料限制而未能納入模型當中。未來若資料可取得，建議一同納入研究。
6. ADAS 警示事件可更多樣。本研究利用 ADAS 警示事件測量事故風險，因為觸發 ADAS 警示的行為往往就是常見的事故原因，若有更多樣的駕駛行為方面 ADAS 警示事件，便能涵蓋到更多樣事故原因。常見之事故原因不只有本研究考量到的未保持安全距離、變換車道不當、超速，還有未依規定讓車、轉彎不當等等。
7. 風險群可劃分成更多群。分成兩群以上，雖容易面臨風險群命名上的困難，但若存在特定數值趨勢，此時仍建議劃分成兩群以上，以更深入刻劃不同風險群的特性。劃分的群數越多，在不同 ADAS 警示事件上的數值相對高低排名，更難相同。但分成兩群以上時，仍有可能觀察到特殊的樣態，雖然實例測試並無觀察到。比如群一於前方警示，如前方碰撞、車距過近等，明顯偏高；群二於側方警示，如車距偏移、側方間隔過近，明顯偏高；群三則於各警示次數數皆較低。此時可以命名群一為前方相對高風險群；群二為側方相對高風險群；群三則為相對低風險群。
8. 車速車距資料品質可更好。本研究使用之車速車距資料存在一些異常值，比如在市區道路時速 200，或者車距為負數，可見資料品質有進步空間。
9. 分析不同道路等級的資料。本研究的道路等級研究範圍為一般道路，而摒除國道上的資料。建議未來的研究可以嘗試分析不同的道路等級，比如只分析國道資料或者只分析一般道路中的郊區道路等等。
10. 由於不同變數挑選方法之結果不相同，建議未來進行變數挑選時，同時參考多個方法的結果，避免結果偏頗。
11. 在實際輸入資料之前，難以確定何模型最佳，故建議先建立多個模型，再用模型品質衡量指標，如 AUC，挑選最終模型。

參考文獻

- [1] 內政部警政署統計查詢網. (2022). 事故概況統計.
- [2] 交通部統計查詢網. (2022). 道安會統計.
- [3] 周榮昌, 邱裕鈞, 郭仲偉, 王明智, 陳孜穎, 謝志偉, 張開國, 葉祖宏與陳凱斌. (2019). 道路交通事故成本推估之研究. 交通部運輸研究所合作研究計畫.
- [4] 林豐福, 張開國與葉祖宏. (2003). 機車駕駛行為與事故風險分析. 交通部運輸研究所合作研究計畫.
- [5] 溫基信, 王宏生, 蘇志文與蔡洛緯. (2019). 道路交通車流及事故風險偵測與分析工具之發展應用. 交通部運輸研究所合作研究計畫.
- [6] Abdel-Aty, M., Cai, Q., Wu, Y., & Zheng, O. (2022). Evaluation of automated emergency braking system's avoidance of pedestrian crashes at intersections under occluded conditions within a virtual simulator. *Accident Analysis and Prevention*, 176, 106797–106797.
- [7] Aurenice, C., & Ana, P. (2020). Proposal of a driver profile classification in relation to risk level in overtaking maneuvers, *Transportation Research Part F: Traffic Psychology and Behaviour*, 74, 375-385.
- [8] Chen, H., Feng, S., Pei, X., Zhang, Z., & Yao, D. (2017). Dangerous Driving Behavior Recognition and Prevention Using an Autoregressive Time-Series Model. *Tsinghua Science and Technology*, 22(6), 682–690.
- [9] Dean, M. E., Lubbe, N., Fredriksson, R., Sternlund, S., & Gabler, H. C. (2023). Assessing the applicability of impact speed injury risk curves based on US data to defining safe speeds in the US and Sweden. *Accident Analysis and Prevention*, 190, 107151–107151.
- [10] Feng, G., & Youjia, F. (2013). Individual driver risk assessment using naturalistic driving data. *Accident Analysis & Prevention*, 61, 3-9.
- [11] Gulino, M., Gangi, L. D., Sortino, A., & Vangi, D. (2021). Injury risk assessment based on pre-crash variables: The role of closing velocity and impact eccentricity. *Accident Analysis and Prevention*, 150, 105864–105864.
- [12] Kidd, D. G., & McCartt, A. T. (2016). Differences in glance behavior between drivers using a rearview camera, parking sensor system, both technologies, or no technology during low-speed parking maneuvers. *Accident Analysis and Prevention*, 87, 92–101.

- [13] Masello, L., Sheehan, B., Castignani, G., Shannon, D., & Murphy, F. (2023). On the impact of advanced driver assistance systems on driving distraction and risky behaviour: An empirical analysis of Irish commercial drivers. *Accident Analysis and Prevention*, 183, 106969–106969.
- [14] Max, K., & Johnson, K. (2019). *Feature Engineering and Selection: A Practical Approach for Predictive Models* (1st ed.). Chapman and Hall/CRC.
- [15] Mehdizadeh, A., Alamdar Yazdi, M. A., Cai, M., Hu, Q., Vinel, A., Rigdon, S. E., Davis, K., & Megahed, F. M. (2021). Predicting unsafe driving risk among commercial truck drivers using machine learning: Lessons learned from the surveillance of 20 million driving miles. *Accident Analysis and Prevention*, 159, 106285–106285.
- [16] Nidamanuri, J., Nibhanupudi, C., Assfalg, R., & Venkataraman, H. (2022). A Progressive Review: Emerging Technologies for ADAS Driven Solutions. *IEEE Transactions on Intelligent Vehicles*, 7(2), 326–341.
- [17] Ospina-Mateus, H., Quintana Jiménez, L., & López-Valdés, F. J. (2023). Analyzing traffic conflicts and the behavior of motorcyclists at unsignalized three-legged and four-legged intersections in Cartagena, Colombia. *Accident Analysis and Prevention*, 191, 107222–107222.
- [18] Tanishita, M., & Van, W. B. (2017). Impact of vehicle speeds and changes in mean speeds on per vehicle-kilometer traffic accident rates in Japan. *IATSS Research*, 41(3), 107–112.
- [19] Yu, R., Han, L., & Zhang, H. (2021). Trajectory data based freeway high-risk events prediction and its influencing factors analyses. *Accident Analysis and Prevention*, 154, 106085–106085.
- [20] Wang, X., Zhou, Q., Quddus, M., Fan, T., & Fang, S. (2018). Speed, speed variation and crash relationships for urban arterials. *Accident Analysis and Prevention*, 113, 236–243.
- [21] Wang, J., Huang, H., Li, Y., Zhou, H., Liu, J., & Xu, Q. (2020). Driving risk assessment based on naturalistic driving study and driver attitude questionnaire analysis. *Accident Analysis and Prevention*, 145, 105680–105680.
- [22] Wickens, C., Mann, R., Stoduto, G., Butters, J., Ialomiteanu, A., & Smart, R. (2012). Does gender moderate the relationship between driver aggression and its risk factors? *Accident Analysis & Prevention*, 45, 10-18.
- [23] Wu, K. F., & Wang, L. (2021). Exploring the combined effects of driving situations on freeway rear-end crash risk using naturalistic driving study data. *Accident Analysis and Prevention*, 150, 105866–105866.
- [24] Xiong, X., Wang, M., Cai, Y., Chen, L., Farah, H., & Hagenzieker, M. (2019). A forward collision avoidance algorithm based on driver braking behavior. *Accident Analysis and*

Prevention, 129, 30–43.

- [25] Xu, C., Louw, T. L., Merat, N., Li, P., Hu, M., & Li, Y. (2023). Drivers' gaze patterns when resuming control with a head-up-display: Effects of automation level and time budget. *Accident Analysis and Prevention, 180*, 106905–106905.
- [26] Zhang, X., Wang, X., Bao, Y., & Zhu, X. (2022). Safety assessment of trucks based on GPS and in-vehicle monitoring data. *Accident Analysis and Prevention, 168*, 106619–106619.
- [27] Zhao, X., Lu, T., Dai, Y. & Omar, C.(2021). Individual Driver Crash Risk Classification Based on IoV Data and Offline Consumer Behavior Data. *Mobile Information Systems, 2021*, 1–10.
- [28] Ziebinski, A., Cupek, R., Grzechca, D., & Chruszczyk, L. (2017). Review of advanced driver assistance systems (ADAS). *AIP Conference Proceedings, 1906*(1).

