CONSIDERING MICROSCOPIC BEHAVIOUR OF MOTORISTS IN SAFETY ASSESSMENT FOR RAILWAY LEVEL CROSSING IN AUSTRALIA

Seunghyeon Lee, Ph.D., Corresponding Author
Assistant Professor
Department of Transportation Engineering
University of Seoul, Seoul, Korea, 02504
Email: seunghyeon.lee@uos.ac.kr

Tuo Mao, Ph.D.
Faculty of Engineering and IT
University of Technology Sydney, Ultimo NSW 2007, Australia
Email: tuo.mao@uts.edu.au

Yuming Ou, Ph.D.
Faculty of Engineering and IT
University of Technology Sydney, Ultimo NSW 2007, Australia
Email: yuming.ou@uts.edu.au

Adriana-Simona Mihaita, Ph.D.
Faculty of Engineering and IT
University of Technology Sydney, Ultimo NSW 2007, Australia
Email: adriana-simona.mihaita@uts.edu.au

Fang Chen, Ph.D.
Faculty of Engineering and IT
University of Technology Sydney, Ultimo NSW 2007, Australia
Email: fang.chen@uts.edu.au

Word Count: 6488 words + 4 table(s) × 250 = 7488 words

Submission Date: July 30, 2022

ABSTRACT

This work proposes a methodology considering the dynamic features of motorists in the vicinity of railway level crossings (RLCs) to the Australian Level Crossing Assessment Model (ALCAM). The dynamic features include real-time train movements, the severity of traffic incidents nearby RLCs, and the risky manoeuvres of motorists when crossing RLCs. We collected train timetables and network data provided by the TfNSW Open Data Hub and real-time trajectories of trains provided by GTFS feeds (General Transit Feed specification) to demonstrate real-time train movements around level crossings. To quantify the severity of traffic incidents in the RLC vicinity, we used NSW road crash data sets, including crash, number of victims, and unit profiles. To identify driving risky manoeuvres, we investigated the speed, XY acceleration, and the G-force of road vehicles to define safe-points and brake-points in the vicinity of RLCs. Furthermore, we employed a neural network clustering method to identify dynamic characteristics of each safe-point and brake-point in the vicinity of RLCs. We integrated three dynamic elements to the nominated weightings with sensitivity analysis. The proposed method lays the foundation stone for considering dynamic features of motorists and assess the rail crossing safety by using public open data. It paves the way for analysing the historical behaviour of motorists in the vicinity of RLCs and anticipate key risks at each RLC in Australia.

Keywords: Safety assessment, Railway level crossing, Behaviour of motorists, Neural network clustering

INTRODUCTION

In Australia, securing safety at railway level crossings (RLCs), where motorists and trains intersect at the same grade, is one of the crucial concerns in an integrated multi-modal transport network. Crashes at RLCs have involved severe fatalities, injuries, and came with significant economic expenses even if the number of crashes at the RLC constitutes below 1% of total crashes in New South Wales (NSW) in Australia for the last several decades. For example, Salmon et al. (*1*) analysed the systemic and psychological factors causing the tragedy at an RLC near Kerang, Victoria, Australia, on 5 June 2007. Eleven train passengers were killed by a loaded semi-trailer truck striking a passenger train at the RLC. This incident additionally involved 13 injured passengers and an injured truck driver. They used two juxtaposed human factors approaches to produce insights into the contributory factors underlying the incident. They concluded that the tragic event could be caused by the lack of fully active controls from the systemic perspective and a Looked-But-Failed-To-See (LBFS) error triggered by an inappropriate schema from the psychological perspective.

To systematically assess the vulnerability of an individual RLC location preventing prevent recurrence of the tragedy in Australia, the national ALCAM group developed the Australian Level Crossing Assessment Model (ALCAM), which is an assessment tool used to identify key potential risks at level crossings and to assist in the prioritisation of crossings for upgrades in (*2*). This complex scoring process is outlined in Australian Standard and New Zealand Standard (AS/NZS 4360:2004) includes a matrix of weightings that measures the influence of the designated characteristics at RLCs on the possible accident mechanisms. ALCAM continues to be updated with fine-tuning weightings to consider new control technology and modifications. Due to its reliable and robust assessment mechanism for RLCs, the Australian Transport Council of Commonwealth, State and Territory, the Standing Committee of Transport (SCOT), all state and territory transport ministers, and New Zealand agreed to adopt the ALCAM to identify potential accident causal factors and overall effects of proposed treatments.

This study aims to introduce the dynamic characteristics of motorists and trains in addition to the factors already involved in the ALCAM. We used a data-driven analysis based on NSW open-sourced data to identify hot spots of RLCs in NSW, Australia. In this study, the dynamic characteristics include real-time train movements, the severity of traffic incidents nearby RLCs, and risky manoeuvres of motorists nearby RLCs. We collected train timetables and network data from TfNSW Open Data Hub to explain the real-time train movements at an individual RLC location. We used NSW road crash data sets, including crash, persons, and unit profiles, to estimate the severity of traffic incidents nearby each RLC. We investigated the acceleration, deceleration, G-force, and speed of road vehicles to understand risky manoeuvres of motorists defined by safe-points and brake-points in the vicinity of RLCs. We identified hot spots of RLCs in NSW in the proposed assessment method by integrating the three dynamic elements to the nominated weightings with a sensitivity analysis. The specific contributions of the study are given below:

- investigating the microscopic dynamics of motorists and the severity of traffic incidents enables to introduce **near misses** in the vicinity of RLCs in a safety assessment mechanism.
- the data-driven analysis based on open-sourced data takes full advantage of **data accessibility and flexibility** if a large amount of data is secured.
- **the sensitivity analysis** helps to understand the varied influence of each dynamic factor on the identification of hotspots of RLCs.

To achieve the research goal, this article is organised as follows. The literature review for safety assessment methods for RLCs and data profiles is presented in Sections 2 and 3, respectively. The assessment methods and results for dynamics of motorists and trains nearby RLCs are demonstrated in Section 4. Finally, Section 6 provides the conclusions of this study and future research directions.

LITERATURE REVIEW

This section summarises the studies on railway level crossings, considering the behaviour of motorists and trains in an RLC safety assessment and identifies the research gaps this study tackled.

Behaviour of motorists in the vicinity of RLCs

RLCs have been regarded as hotspots of reduced traffic safety in multi-modal urban networks since a single crash between trains and vehicles has brought about its swingeing damage at RLCs. Understanding behaviour of motorists in the vicinity of RLCs is one of the key factors to secure safe operating and driving environment at RLCs in (*3*), (*4*), (*5*), (*6*), (*7*), and (*8*).

Tenkink and Van der Horst (*4*) examined the behaviour of car drivers at two RLCs, in which flashing warning lights are installed using analysing recorded video. Evans (*9*) investigated fatal accidents and fatalities at RLCs in Great Britain from 1946 to 2009 from a macroscopic perspective. They found that the number of fatalities per year have decreased by 65% in the first half of the study period, whereas after that they have been relatively constant by the end of the study period. Evans and Hughes (*10*) continued to conduct the safety investigation at RLCs from a microscopic perspective in Great Britain. They investigated relationships between traverses, delays, and fatalities to road users at RLCs to confirm an additional risk for the users.

Salmon et al. (*8*) mentioned that the situation awareness and the decision-making of motorists are not sufficiently considered in existing methodologies to evaluate the safety performance of RLCs. They conducted an on-road network analysis to examine the situation awareness of motorists at RLCs. Zhao and Khattak (*11*) investigated injuries of motorists involved in train-motor vehicle crashes at RLCs using three injury severity models based on ordered probit, multinomial logit, and random parameter logit. They found that the likelihood of more severe crashes increases when they involve higher train and vehicle speeds, freight trains, older drivers, and female drivers. Liang et al. (*12*) analysed the macroscopic behaviour of motorists while crossing RLCs during the closure cycle. They used the violation rate of speeding to define the risky behaviour of vehicle drivers estimated during RLCs closure cycles.

Safety assessment for RLCs

Gitelman and Hakkert (*13*) used a hazard index to evaluate the RLCs safety based on limited historical accident statistics in Israel. Russell and Mutabazi (*14*) adopted to calculate a designated rating for each considered feature to construct a ranking table for RLCs closures in Kansas, the United States. Miranda-Moreno et al. (*15*) evaluated Canadian RLCs based on an accident history of 5 years using the traditional negative binomial model, the heterogeneous negative binomial model, and the Poisson log-normal model. They confirmed that a list of hazardous RLCs significantly depends on the model assumptions and raking criteria. Hu and Wu (*16*) conducted an empirical study to identify accident frequency and severity. They employed a logit model to categorise severity into three levels: no-severity, minor-severity, and serious-severity. Gruyter and Currie (*17*) carried out a detailed literature review to develop an international synthesis of RLCs impacts. They revealed 18 different types of impacts associated with RLCs; meanwhile, safety defined by accident frequency and severity at RLCs is one of the essential factors in an RLC safety assessment process. They asserted that using empirical evidence to support impact assessments is required to understand RLCs impacts better.

Recently, a machine learning approach was introduced for safety assessment for RLCs. Soleimani et al. (*18*) introduced a machine learning approach to develop a comprehensive RLC Consolidation Model. They mainly considered train operational factors, geometric conditions of RLCs, and traffic conditions at RLCs. The developed classifier based on the extreme gradient boosting (XGboost) algorithm produced that 62% of current RLCs should be closed or improved in Louisiana, the United States. Haleem (*19*) analysed safety issues and improvement measurements at private RLCs in the United States. To identify safety problems linked to increased injuries and fatalities, they used temporal crash characteristics, geometry, railroad, traffic, vehicle, and environment predictors in mixed logit models. Keramati et al. (*20*) used the random survival forest (RSF) to investigate the crash severity at RLCs. They found that additional stop signs and audible devices could reduce the crash likelihood, property damage only, injuries, and fatal crashes at RLCs. Keramati et al. (*21*) analysed the influence of geometric conditions, including acute crossing angle, the number of tracks, the distance from the RLC to the nearest intersection, and the number of lanes, on crash occurrence and severity likelihoods at RLCs.

Research gaps

We discover two significant research gaps from the comprehensive review studies to introduce dynamic characteristics of motorists and trains in the vicinity of RLCs in a safety assessment process.

First of all, current safety assessment processes of RLCs mainly deal with geometric conditions of RLCs, road and railway traffic volumes, and historical crash records. In the meantime, microscopic dynamics of motorists have been overlooked in the safety assessment processes of RLCs due to difficulties in data collection and unclear linkages between microscopic dynamics of motorists and safety at RLCs. Over the past several decades, a variety of methods have been developed to model longitudinal and lateral interactions of vehicular movements on the road, called a car-following (CF) and a lane-changing (LC) model, respectively, at the microscopic level in (*22*). The modelling of these two microscopic two-dimensional manoeuvres has played a significant role in the comprehensive understanding of characteristics of traffic flows as well as traffic safety in an urban network. In this study, we investigated acceleration, deceleration, G-force, and speed of motorists to analyse near misses in the vicinity of RLCs from the perspective of a safety assessment of RLCs.

Second, open-sourced data sets were not attractive options in safety assessment methods of RLCs due to data quality, reliability, and accessibility. In recent years, the debut of information and communication technologies (ICT) in traffic engineering has diversified spatial-temporal dimensions and resolutions of available data sets for transport systems analysis. A multitude of traffic data sources is available on an urban transportation network to improve temporal and spatial coverages of intermodal traffic patterns in real-time. The multiple data-sources include fixed sensors and mobility sensors in (*23*), (*24*), (*22*) and (*25*). Conventional fixed sensors are a relatively reliable and robust-data source, whereas expensive construction and maintenance costs are inherent in this type of sensor. In the meantime, personal mobility data sources can provide the detailed spatial and temporal behaviour of moving objects' movements, although this type of data requires effective statistical sampling techniques because of the sparsity of available data. In this study, we used train timetables and crash data sets obtained from TfNSW Open Data Hub and FCD derived from GPS data from Compass IoT in NSW to maximise data accessibility. Moreover, we conducted the sensitivity analysis among safety features at RLCs to understand the varied influence of each dynamic factor on the identification of hotspots of RLCs and maximise the flexibility of data analysis.

DATA ANALYSIS

We secured five data sets from four different data sources. The first data is an information sheet about RLCs obtained from NSW public level crossing finder[1]. The information sheet contains the profiles of 1360 RLCs, including identification number, road name, control type, owner, line section, suburb, number of tracks, longitude, and latitude.

The second data is a historical unplanned incident profile from 16 August 2006 to 11 August 2021 obtained from Traffic Management Centre (TMC) in NSW. The historical incident profile involves over a million incident records that have been recorded together with information regarding location, duration, type, link ID, segment ID, direction, and Sector ID. These historical incident records allowed us to access high-resolution location information of a single incident for hot-spots identification of RLCs; meanwhile, the severity of crashes was not provided.

The third data is a historical crash profile from 1 January 2015 to 31 December 2019 provided by TfNSW Open Data Hub[2]. This data set contains crash profiles and information on the people and units involved. In crash profiles, they provided a degree of the crash, street of the crash, type of location, primary permanent, temporary, and hazardous features, surface conditions, natural lights, signals operations, speed limit, users movement, short description, first impact type, the number of traffic units involved, and the number of killed, the number of serious, moderately, and minor-other injured. They linked crash profiles to profiles of people involved, including gender, age group, road user class, degree of casualty, and units, containing a role in the first impact, types, direction of travel, manoeuvre and the first and the second crashed objects.

The fourth data consists of railway timetables and network profiles from 1 January 2019 to 31 August 2021 obtained from TfNSW Open Data Hub and Google GTFS. Railway timetables provide information on a train service schedule, including departure station, departure time, arrival station and arrival time for each passenger train service. The Railway network includes information on train line geometry and train station location.

The fifth data is microscopic dynamics of motorists from 1 March 2021 to 31 August 2021 provided by Compass IoT [3]. Compass IoT is a data aggregation and analytic company that uses connected vehicles to generate in-depth insights into vehicular trajectories across transport networks. They collected trajectories data from private and public data providers to provide the average speed profiles and volumes in a single link across the Australian road network. The collected microscopic dynamics of motorists enable us to evaluate the driving behaviour of vehicles in the vicinity of RLCs by analysing vehicular manoeuvres, including braking, acceleration, and deceleration of a single-vehicle.

These five data sets were used to analyse profiles of RLCs, traffic incidents, railway operations, and microscopic behaviour of motorists described in the following subsections.

[1] https://appln.transport.nsw.gov.au/mapservices/proxy/levelCrossings/map.html
[2] https://opendata.transport.nsw.gov.au/
[3] https://console.compassiot.cloud/

Railway level crossing

We investigated general information on RLCs in NSW. The data set of train level crossings in NSW contains the following features: CROSSING NUMBER, ROAD NAME, LOCATION, CONTROL TYPE, CONTROL CATEGORY, RAIL OWNER, LINE SECTION, LGA NAME, SUBURB, RAIL KM, NO OF TRACKS, OBJECT ID, lon, and lat. The features of lon and lat were used for map-matching in a QGIS program.

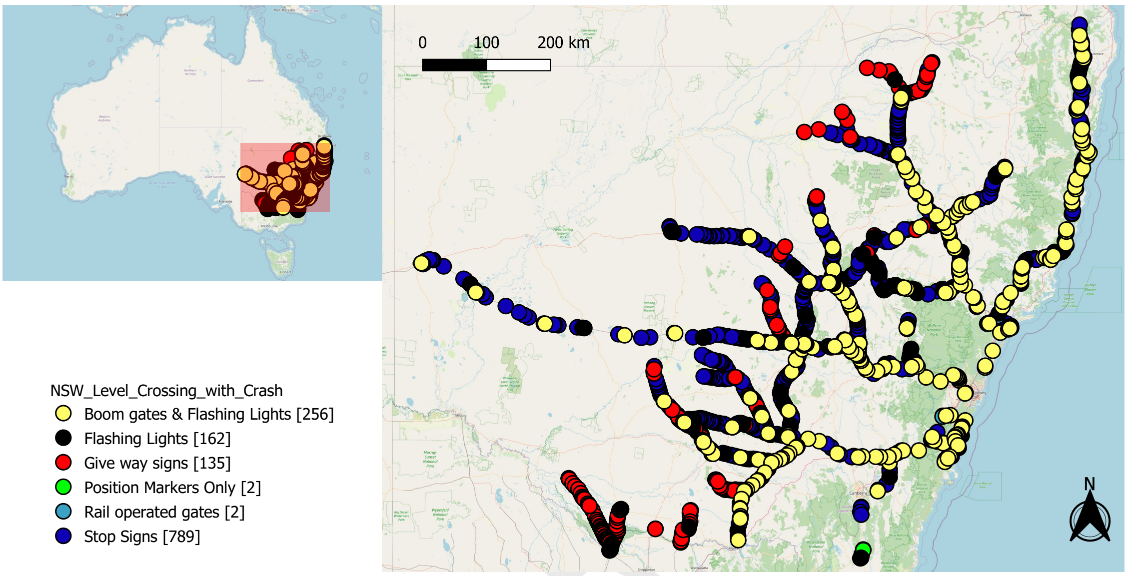


FIGURE 1: The distributed locations of railway level crossings in NSW.

Active and passive types of level crossing controls are installed across NSW. In the active type, boom gates and flashing lights are installed at 256 RLCs. Flashing lights are operated at 162 of RLCs across NSW. Rail-operated gates are installed at two places of RLCs. In the passive type, stop signs and give way signs are installed at 789 and 135 RLCs, respectively. Fig. 1 presents distributed locations of each level crossing control category at RLCs. Stop signs in the passive type are widely installed across regional NSW, whereas boom gates and flashing lights in the active type are mainly installed in the Sydney metropolitan area.

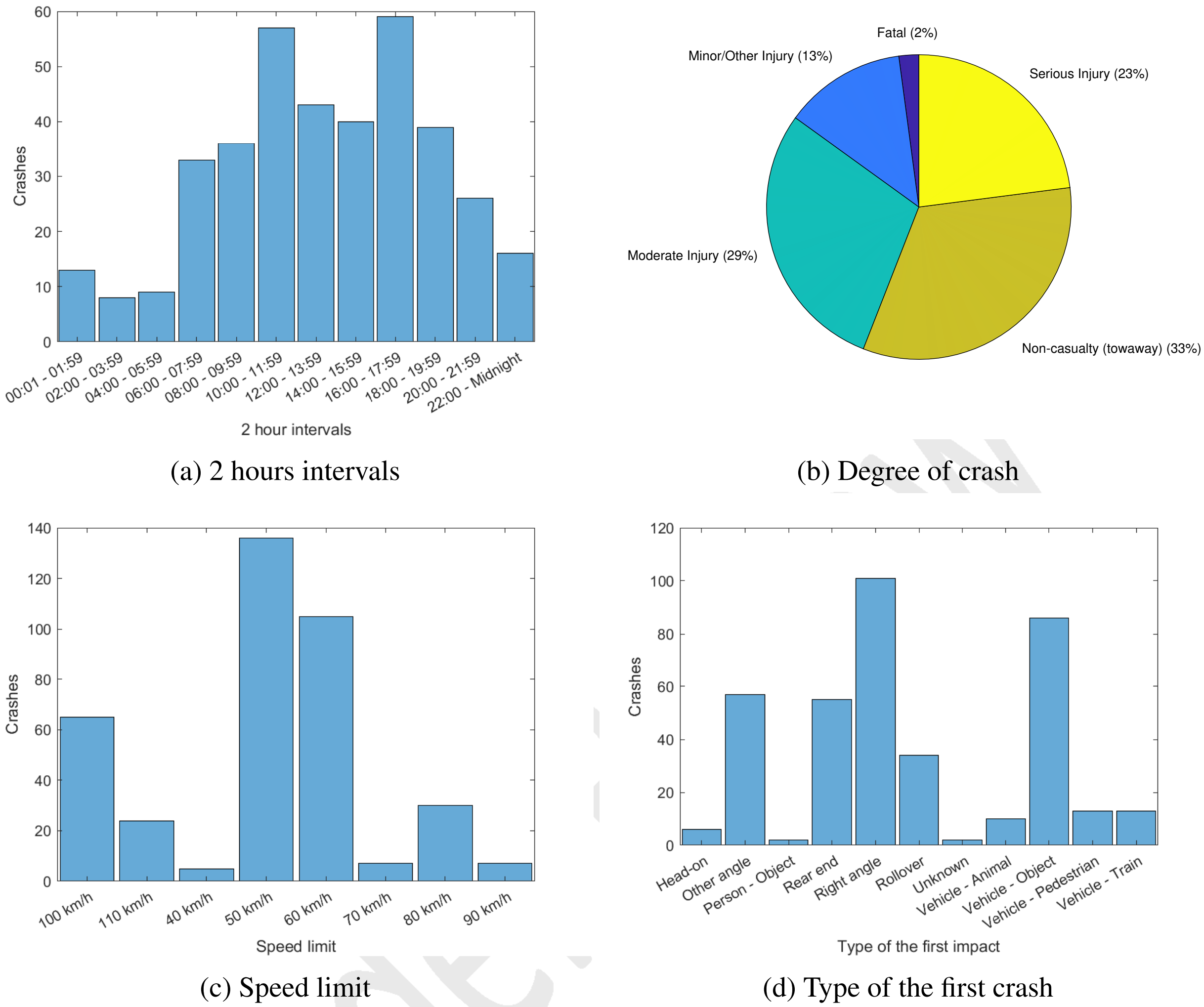
Traffic incidents

We investigated two open-sourced traffic incident data sets, including historical unplanned incident profiles over ten years and crash profiles for five years. Unplanned incident logs present the summarised information of over a million incidents in NSW, whereas crash profiles illustrate the detailed information of each crash of over 100 thousand. In unplanned incident logs, we filtered this data set to extract only train-related incidents, and the remaining findings revealed that 2,549 were train-related incidents. The data revealed an average duration of 139 minutes or train disruption, representing quite a considerable time interval with no train service in affected areas or trains blocked in specific train networks without being able to continue their journey. The maximum duration of a train incident has reached 7,663 minutes and was signalled in Glenyalla north of Kankool on 15 February 2015.

In crash profile data, we analysed crashes, units involved in crashes, and people involved in crashes in the vicinity of RLCs. We filtered 379 crashes nearby RLCs among all crashes from

1 January 2015 to 31 December 2019. The definition of the vicinity of RLCs is within a radius of 150m from the location of RLCs in this study. In Fig. 2, we illustrated characteristics of crashes in the vicinity of RLCs. Most crashes have occurred from 6:00 to 19:59 in Fig. 2a. Two-time intervals, 10:00 – 11:59 and 16:00 – 17:59, include over 100 crashes for the last five years. We found around a one-hour time lag between peak hours from the perspective of road traffic and peak intervals of crashes nearby RLCs. 67% of crashes in the vicinity of RLCs have included injured people in Fig. 2b. Moreover, 23% of crashes have contained seriously injured people, in which this proportion is much more significant than that involved in general road crashes. Although most crashes were on the road with 50 and 60 km/h of the speed limit, over 100 crashes have occurred on the road over 80 km/h of the speed limit in Fig. 2c. Almost 200 crashes have occurred nearby RLCs where the control device is not installed. Moreover, active control devices at RLCs have played a critical role in preventing crashes nearby RLCs. In Fig. 2d, the majority of crashes are involved right-angle crashes and vehicle-to-object crashes. Vehicle-to-train direct crashes occurred below 20 crashes over the last five years. One hundred fifty crashes occurred on the two-way undivided road, whereas T-junction and X-intersection included around 100 crashes each nearby RLCs.

We found three significance from analysing RLCs statistics. First of all, crashes in the vicinity of RLCs have generally involved severe injuries in NSW (Fig. 2b. In addition, we found time lags between peak hours of road traffic and peak time intervals of crashes nearby RLCs in NSW (Fig. 2a). Furthermore, analysing all crashes nearby RLCs helps understand safety issues at RLCs because the number of direct crashes between vehicles and trains is small among first crashes nearby RLCs in NSW (Fig. 2d).



(a) 2 hours intervals

(b) Degree of crash

(c) Speed limit

(d) Type of the first crash

FIGURE 2: Crashes in the vicinity of RLCs

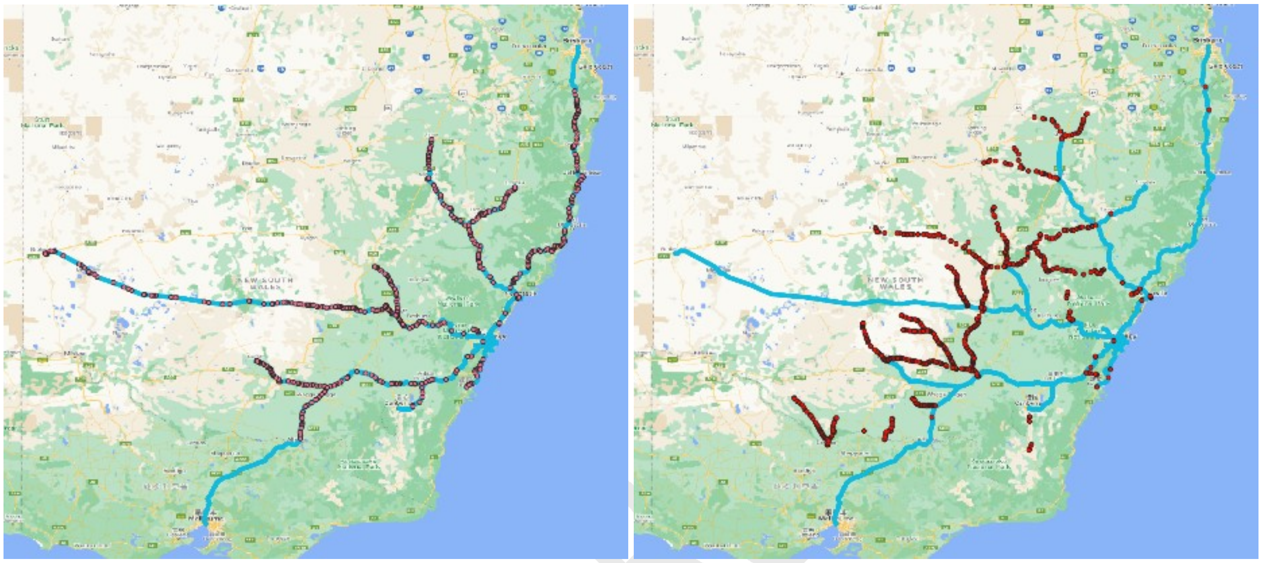
Railway operations

This section describes the frequency of passenger trains in NSW using RLC train timetable data. The service frequency of railway systems reflects how busy the specific RLC is in terms of trains passing the RLC. The train frequency has been an essential characteristic of RLCs in developing safety scores at RLCs in ALCAM for the last several decades. The busier RLC has increased the risk of crashes and more severe consequences.

We used three data sets: RLC profiles, train timetables, and train networks. RLC profiles collected from NSW Public Crossing Finder contain ID, road name, control type, owner, line section, suburb, number of tracks, longitude, and latitude. Train timetables involve departure station, departure time, arrival station and arrival time for each passenger train service. The data was collected from TfNSW Open Data Hub. Train network data include geometric conditions of train lines and locations of train stations. The data was also collected from TfNSW Open Data Hub.

Three data sets were processed through two steps: map matching and arrival time estimating. In the map matching, the locations of the RLCs are not precisely located on the train lines since the RLCs data and the train network data set are collected from different organisations. As a result, we employed a map matching technique to find the correct location of the RLCs on the train lines.

We identified 590 (43%) of RLCs, which are operating with the given timetables in the passenger train line, while 770 (57%) of RLCs are not. They are illustrated in Fig. 3a and Fig. 3b, respectively. In the estimating step of train arrival time at RLCs, we excluded 770 of RLCs from the investigation because the given passenger timetable data do not cover them.



(a) RLCs that can be projected

(b) RLCs that cannot be projected

FIGURE 3: The results of the map matching

After the map matching of the RLCs, we estimated the arrival time of each train service at each RLC. We used the departure time at the last train station and the arrival time at the next train station to estimate the RLC arrival time. As shown in Fig. 4, there is a level crossing L, which is located between two stations S_1 and S_2.

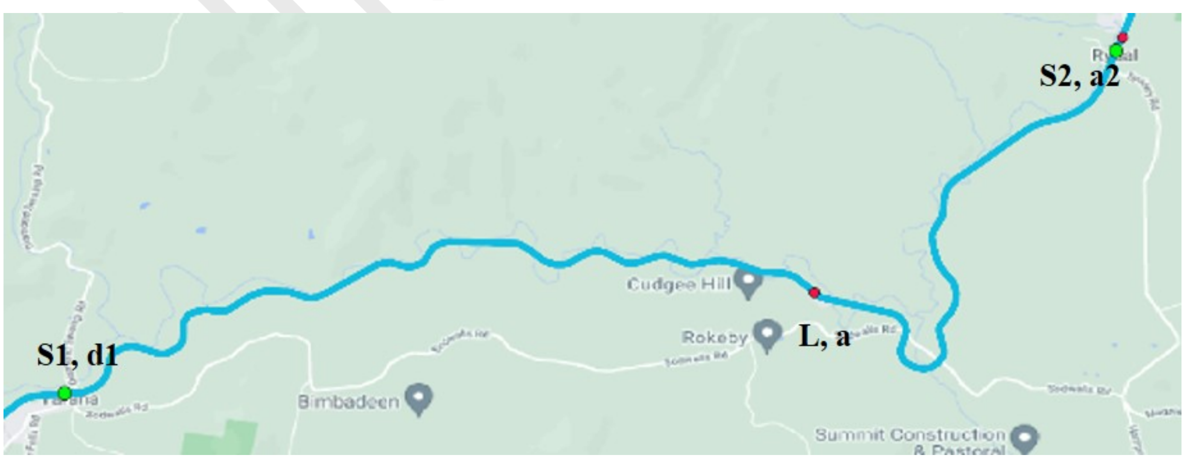


FIGURE 4: Estimating arrival time at RLCs.

The arrival time at the RLC, L, is estimated by using the following equation:

$$a = d_1 + \frac{Distance_{S_1,L} \times (a_2 - d_1)}{Distance_{S_1,S_2}} \tag{1}$$

where d_1 is the departure time at station S_1,
a_2 is the arrival time at station S_1,
$Distance_{S_1,L}$ is the distance from station S_1 to RLC L along the train line,
and $Distance_{S_1,S_2}$ is the distance from station S_1 to station S_2 along the train line.

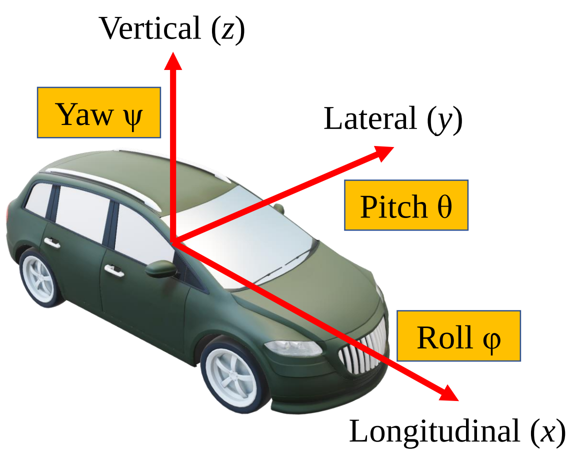
Behaviour of motorists

This section investigated the driving behaviour in the vicinity of RLCs using the Compass IoT data. The primary purpose of this investigation is to identify repetitive risky driving behaviour nearby TLCs with a high risk for collisions or near-misses due to driver misbehaviour.

Compass IoT is a data aggregation and analytic company that uses connected vehicles to generate in-depth insights across transport networks. It is a new innovative way of surveying roads, collecting data, and planning cities. Compass IoT collected data from private and public data providers and developed sophisticated algorithms to predict speed and volume across Australian roads. We analysed three pros and cons of Compass IoT data to assess the safety of RLCs in NSW. First of all, it helps to evaluate the driving behaviour in the vicinity of RLCs. In addition, it helps to analyse the breaking/acceleration/deceleration manoeuvres of drivers nearby the RLCs. Moreover, it can reveal near-misses events based on G-force evaluation. In the meantime, the current fleet for collecting the data is limited to 700,000, in which more vehicles have been passing by each RLC regularly that are not captured in Compass IoT. The Compass IoT only collected autonomous vehicle data from March 2020, so no prior information is available for understanding old crashes and events nearby RLCs.

The Compass IoT data consists of safe-point and breakpoint data. safe-point data is designed to leverage the accelerometer readings from the connected vehicles to determine events of interest in the road network. Specifically, these data sets include observations of harsh vehicle movements, and violent events, indicating a near miss, a collision, an incident, or another high G-force event. brake-point data maps the G-force, speed, curvature, and driving violence every second.

We employed a G-force methodology to analyse the behaviour of vehicles nearby each RLC. A G-force is a measure of acceleration. $1g$ is measured as the acceleration resulting from gravity. Gravity is measured in m/sec^2. The value of acceleration is 9.806 m/sec^2. We used a definition of the X/Y/Z axis specified in ISO 8855:1991 to estimate the G-force. Fig. 5 illustrated the x-axis points towards the front of the vehicle, the y-axis towards the left, and the z-axis upwards (right-hand system) with the origin at the most forward point on the centreline of the vehicle for dynamic data measurements.



Driving manoeuvres		Target g-force
Cornering	Mild Left	0.2 – 0.3 G's
	Moderate Left	0.3 – 0.4 G's
	Hard Left	0.5 – 0.6 G's
	Mild Left	0.2 – 0.3 G's
	Moderate Right	0.3 – 0.4 G's
	Hard Right	0.5 – 0.6 G's
Braking	Mild	0.4 – 0.5 G's
	Moderate	0.5 – 0.6 G's
	Hard	0.6 – 0.7 G's
Acceleration from stationary position	Mild	0.2 G's
	Hard	0.3 – 0.4 G's
Turns	Mild Left	0.2 – 0.3 G's
	Moderate Left	0.4 – 0.5 G's
	Hard Left	0.6 – 0.7 G's
	Mild Left	0.2 – 0.3 G's
	Moderate Right	0.4 – 0.5 G's
	Hard Right	0.6 – 0.7 G's

FIGURE 5: The x/y/z axis definition in vehicular manoeuvres.

The severity of the g-force is represented in the platform as harsh acceleration and harsh braking in Fig. 5. Deceleration or braking from 60 km/h to 0 km/h in 3.0 seconds at 0.6 g is considered as harsh braking. Accelerating from 0 km/h to 60 km/h over 0.5 g's is hard acceleration. Swerving or cornering is considered harsh when it is over 0.47 g's.

We first extracted the harsh driving behaviour around 50 metres, 100 metres, and 150 metres from all public level crossing locations from the safe-point data in Compass IoT. The choice for analysing various radii in the vicinity of the TLCs is to understand the level of details that can be captured in the vicinity of RLCs. While 50 metres might be considered a good location, some TLCs might be in the vicinity of more complex road structures; therefore, the need to extend the analysis to a 150m radius. Secondly, we evaluated all RLCs according to the number of harsh driving behaviours and near misses. Thirdly, we analysed the G-force and the maximum speed to identify typical types of harsh driving behaviours around RLCs. Lastly, we inferred the reasons for these harsh driving behaviour and near misses of the top 10 level crossings. We extracted the average speed, volume, and speed limit around the top ten vulnerable RLCs. In addition, we investigated google street view to further identify the possible improvement of the RLC according to the geometry and traffic signs.

Compass IoT data is collected for 17 months, from 1 March 2020 to 31 August 2021. It includes 62, 109, and 172 near-miss events detected at 50, 100, and 150 metres, respectively, within the radius of RLCs in NSW.

ASSESSMENT RESULTS

We assessed all RLCs in NSW from three perspectives, including incident factors, railway operational factors, and behaviour factors of motorists. Moreover, we provided the integrated final risk index score with sensitivity analysis.

Incident factors

We employed the QGIS program to count how many crashes exist in the designated radius from the location of RLCs in NSW. We identified all crashes near the site of RLCs in Fig. 6. One hundred ninety-two of RLCs included more than one crashes nearby the RLCs. We assigned the weight to the degree of a casualty involved in crashes near each RLC. The values of the weight are 5, 4, 3, 2, and 1 for the killed, the seriously injured, the moderately injured, the minor injured, and the traffic units involved. The top ten safety hotspots are illustrated in Table 1.

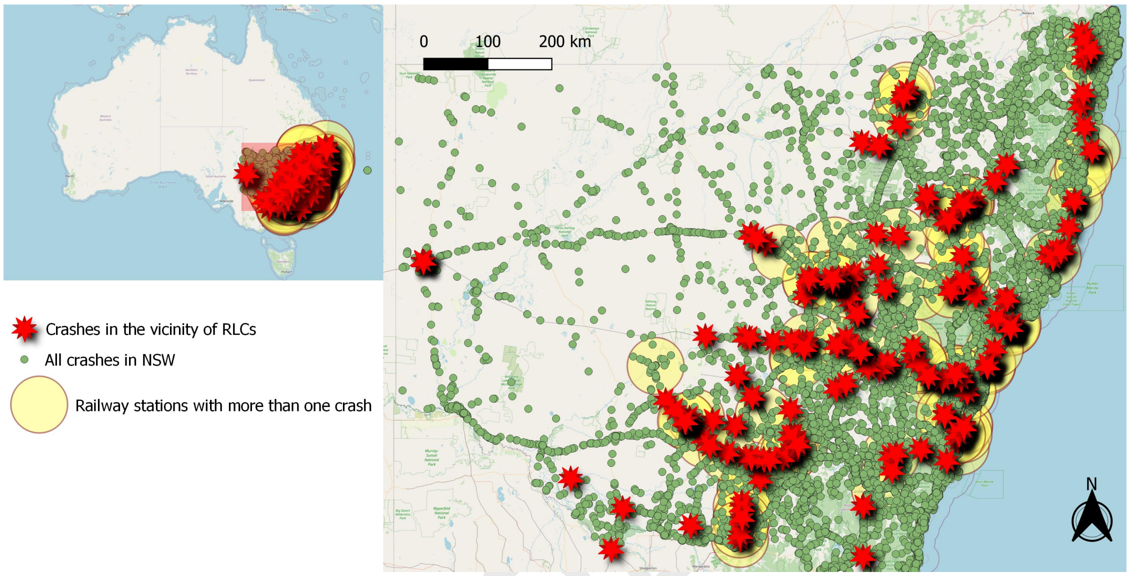


FIGURE 6: Distributed locations of crashes in NSW.

Road name	Control	Suburb	Tracks	Crashes	Safety score	Units involved	Killed	Seriously injured	Moderately injured	Minor injured
Boothenba Road	Boom Gates & Flashing Lights	DUBBO	SINGLE	14	108	29	1	9	10	4
Gosford Road / Rawson Road	Boom Gates & Flashing Lights	WOY WOY	MULTIPLE	16	75	33	0	3	8	3
St James Road	Boom Gates & Flashing Lights	ADAMSTOWN	MULTIPLE	18	72	32	0	4	6	3
Nolan Street	Boom Gates & Flashing Lights	UNANDERRA	SINGLE	6	57	14	0	7	5	0
Bundarra Street	Boom Gates & Flashing Lights	BLACKHEATH	MULTIPLE	10	47	20	0	2	5	2
Summer Street / Mitchell Highway	Boom Gates & Flashing Lights	ORANGE	SINGLE	10	42	16	0	3	4	1
Old Mendooran Road	Stop Signs	DUBBO	SINGLE	5	33	9	0	1	4	4
Darling Street	Boom Gates & Flashing Lights	DUBBO	MULTIPLE	6	30	12	0	0	6	0
Marina Drive	Boom Gates & Flashing Lights	COFFS HARBOUR	SINGLE	4	29	8	0	4	1	1
Brisbane Street	Flashing Lights	EAST TAMWORTH	SINGLE	6	28	12	0	2	2	1

TABLE 1: The safety scores of the top ten hotspots.

In Table 1, the RLC on Boothenba Road produced the highest safety score, 108, due to the involvement of one death, nine seriously injured and ten moderately injured people in the crashes in the vicinity of this RLC. The RLC on St. James Road included the highest number of crashes, 18, with 72 on the safety score.

Railway operational factors

We constructed the ranking tables for entire stations in NSW through railway operational data analysis. In NSW, we observed 57 level crossings with a daily service frequency higher than ten trains per day. The top ten most circulated RLCs in NSW are described in Table 2.

Rank	Road name	Suburb	Frequency (trains/day)
1	Pine Road	Fairfield	241
2	Glenrock Parade	Koolewrong	102
2	Gosford Road	Woy Woy	102
2	Railway Parade	Corrimal	102
2	Bellambi Lane	Bellambi	102
2	Park Road	Woonona	102
7	Beaumont Street	Hamilton	87
8	School Parade	Clifton	84
8	Warnervale Road	Warnervale	84
10	Bandon Road	Vineyard	72
10	Garfield Road	Riverstone	72
10	Mulgrave Road	Mulgrave	72
10	Fairey Road	Windsor	72
10	Bouke Street	East Richmond	72
10	Level Crossing Road	Vineyard	72
10	Racecourse Road	Clarendon	72

TABLE 2: The top ten most circulated RLCs in entire NSW

On average, two hundred forty-one trains per day at the first rank revealed that trains are passing Pine Road every 6 minutes, which might make the road vulnerable RLCs for collisions between trains and motor vehicles. We additionally investigated hourly train service frequency at the RLC on Pine Road. Fig. 7 demonstrates the daily number of trains aggregated per hour. The busiest hour of the day seems to be 7:00 with 15 hours passing by, one train passing by every 4 minutes, followed by 18:00 with 16 trains per day, one train passing by every 3.75 minutes. The RLC on Pine Road is operated by the active boom gates and flashing lights on multiple rail tracks.

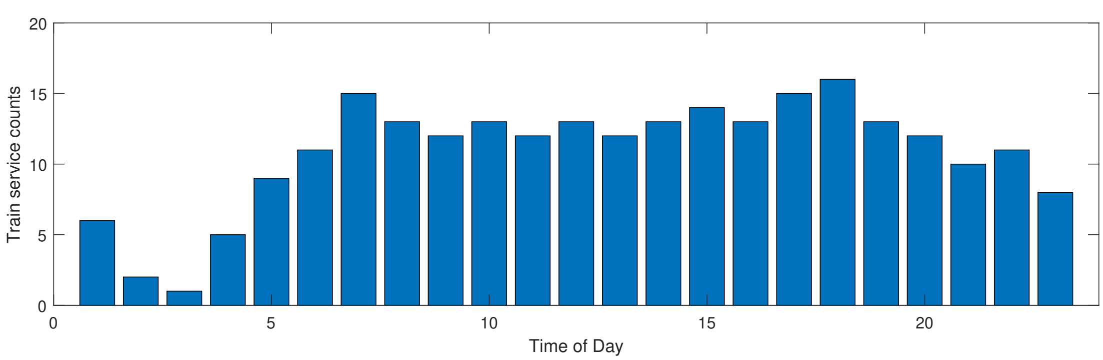


FIGURE 7: Hourly train service frequency at the RLC on Pine Road.

This section investigates the train service frequency at each train level crossing in NSW. We apply a 3-step method, including map matching, arrival time estimation and service frequency counting, to calculate the train service frequencies. Results revealed 57 busy train level crossings in the entire area of NSW and 42 busy train level crossings in the regional area, which are serviced the most by daily train trips. The top-level crossing 945 in the entire NSW has daily 241 train services and more than ten train services every hour in the daytime.

Behaviour factors of motorists

We present the top ten RLCs according to the number of incidents and near misses in a 150-metre radius. In Table 3, all RLCs are controlled by an active type of control. Boom gates and flashing lights control 9 out of 10 RLCs. We detect 18 near misses in the 150-metre vicinity of the RLC on Mulgrave Road, Mulgrave. Furthermore, eight near misses were observed in the vicinity of the RLC on Borenore Road and Amaroo Road, Borenore. Six RLCs involve over five near misses in addition to the above two RLCs across NSW.

Rank	Road name	Suburb	LGA	Control type	Control category	Mear misses
1	Mulgrave Road	Mulgrave	Hawkesbury	Active	Boom gates & Flashing lights	19
2	Borenore Road/ Amaroo Road	Borenore	Cabonne	Active	Boom gates & Flashing lights	8
3	Darling Street	Dubbo	Dubbo	Active	Boom gates & Flashing lights	7
4	Pine Road	Fairfield	Cumberland	Active	Boom gates & Flashing lights	7
5	Fitzroy Street	Dubbo	Dubbo	Active	Boom gates & Flashing lights	6
6	Bandon Road	Vineyard	Blacktown	Active	Boom gates & Flashing lights	5
7	Eulomogo Road	Wongarbon	Dubbo	Active	Boom gates & Flashing lights	5
8	Muldoon Street	Taree	Mid-coast	Active	Boom gates & Flashing lights	5
9	Vitoria Street/ Mitchell Highway	Dubbo	Dubbo	Active	Boom gates & Flashing lights	4
10	Castlereagh Highway	Ben Bullen	Lithgow City	Active	Flashing lights	4

TABLE 3: Top ten level crossings according to near misses in a 150-metre radius.

We analysed the average speed and XY accelerations of all near misses around each RLC in a 150-metre radius. The number of near-misses from the first ranked RLC to the 88th ranked RLC is described in Fig. 8. In the meantime, average speed, average X acceleration, and average Y acceleration are illustrated inFig. 9, Fig. 10, and Fig. 11.

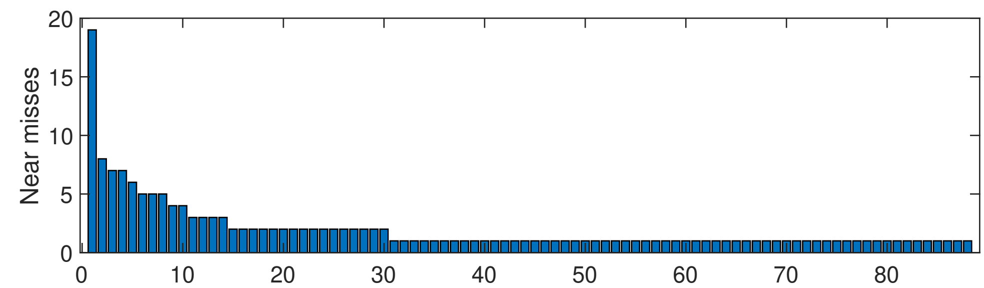


FIGURE 8: Near misses in a 150m radius.

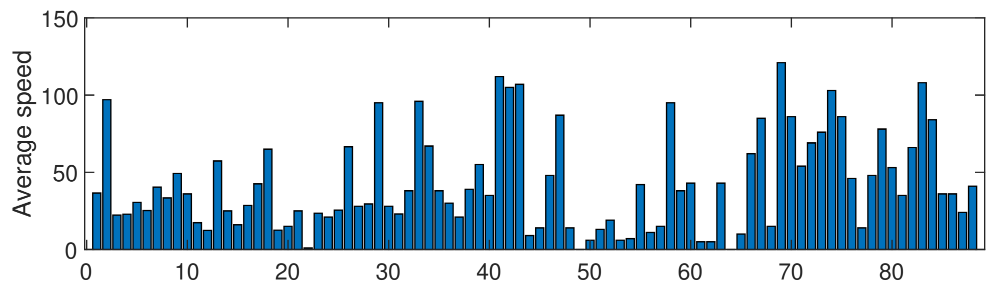


FIGURE 9: Average speed.

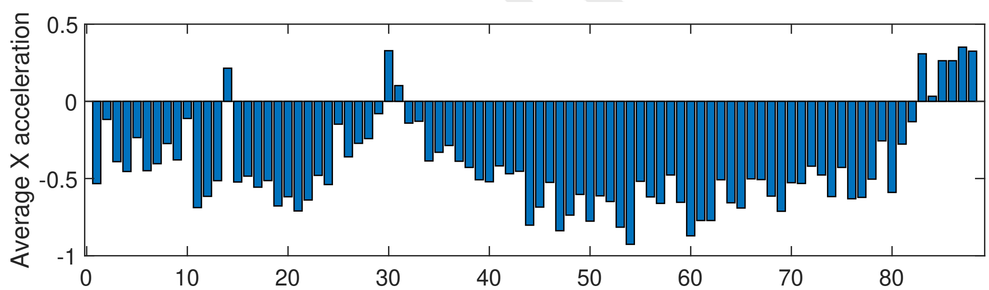


FIGURE 10: Average X acceleration.

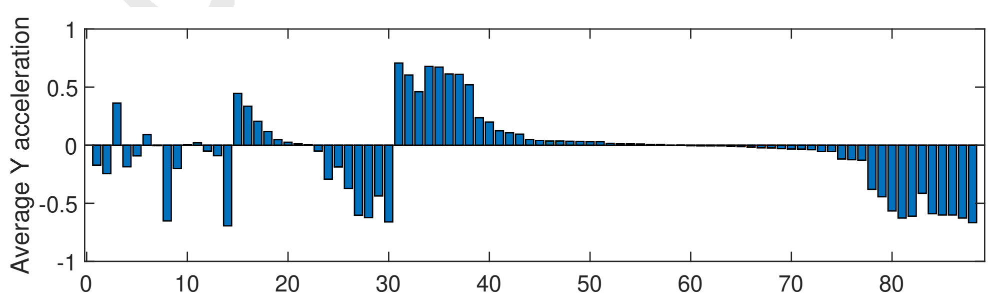
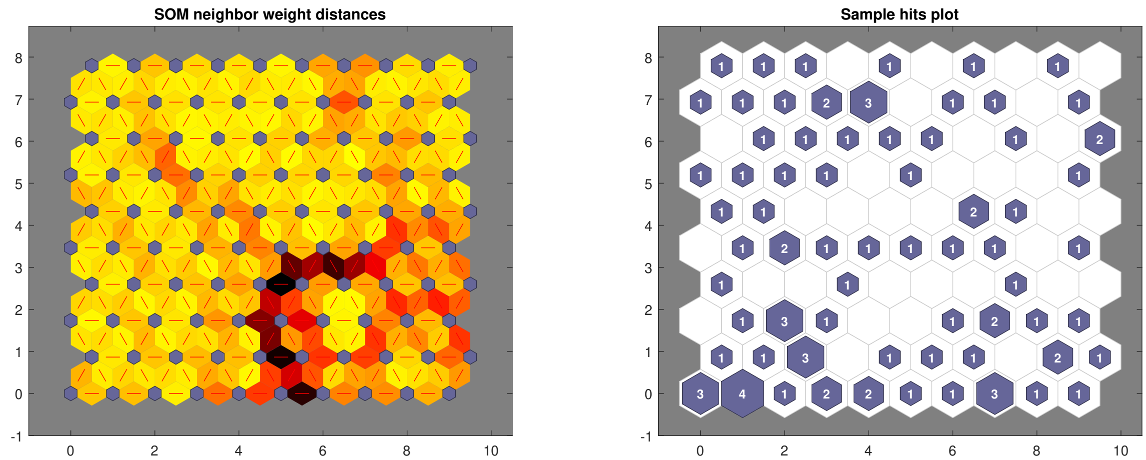


FIGURE 11: Average Y acceleration.

Based on the three dynamic features, we classified RLCs into five groups according to the severity of the g-force. The severity of the g-force is represented in the platform as harsh acceleration and harsh braking in Fig. 5. Deceleration or braking from 60 km/h to 0 km/h in 3.0 seconds at 0.6 g is considered as harsh braking. Accelerating from 0 km/h to 60 km/h over 0.5 g's is hard acceleration. Swerving or cornering is considered harsh when it is over 0.47 g's. The RLCs in NSW are categorised into 18 RLCs as hard brake and stiff steering, 54 RLCs as hard brake and no steering, 4 RLCs as lightweight brake and stiff steering, and 3 RLCs as lightweight brake and no steering, and 9 RLCs as accelerating and stiff steering.

By classifying the behaviours of incidents around RLCs, we also infer the reasons for each type. When drivers do "hard brake and hard steering", they must drive through a curvy road and slow down to avoid an accident harshly. When drivers do "hard brake and no steering", they must drive through a straight road and slow down to avoid an accident harshly. When drivers do "light break and hard steering", they must drive through a curvy road and slow down to avoid an accident slightly. When drivers do "light break and no steering", they must drive through a straight road and slow down to avoid the accident slightly. When drivers do "accelerate and hard steering", they must drive through a curvy road and need to accelerate to avoid an accident. When drivers do "accelerate and have no steering", they must drive through a straight road and need to accelerate to avoid an accident.

Moreover, we adopted a neural network clustering method to propose a novel grouping of vehicular dynamics in the vicinity of RLCs depending on three factors: average speed, average X acceleration, and Y acceleration in the near misses in a 150-metre radius of RLCs. We used the neural network clustering program provided by MATLAB R2022a to group RLCs. In the neural network, we set the map size value to ten, which corresponds to a grid with ten rows and columns, and the map has 100 neurons. We generated self-organising maps (SOM) neighbour weight distances and SOM sample hits to analyse the training results in Fig. 12. Each neuron is represented by each of the hexagons in the plots. The input space is three-dimensions, including three features in each input vector. The weight vectors fall within this space. The relationships among the three-dimensional cluster centres are visualised in two dimensions. In Fig. 12a, the blue hexagons depict the neurons. Red lines in the several coloured regions connect neighbouring neurons. The colours in the regions imply the distances between neurons. The lighter colours (i.e.amber) represent smaller distances, whereas the darker colours (i.e.black) denote more considerable distance. We can classify the RLCs into six or seven divided by a band of darker regions in Fig. 12a.



(a) SOM neighbour weight distances (b) Sample hits plot

FIGURE 12: Self-organising maps with hits

Moreover, we demonstrated how many RLCs are associated with each of the neurons in Fig. 12b. The maximum number of RLCs associated with a neuron is four, in which the neuron is closely related to the surrounding neurons.

Fig. 13 describes a weight plane for average speed, average X acceleration, and average Y acceleration at each RLC. These plots show the weight connecting each attribute to each neuron, with darker colours showing larger weights. According to these plots, three features are not correlated with each other; meanwhile, features play a significant role in clustering RLCs at each platoon of neurons in this analysis. Furthermore, the clustering result by a neural network produced similar grouping results derived from analysing behaviour of incidents around RLCs based on G-force.

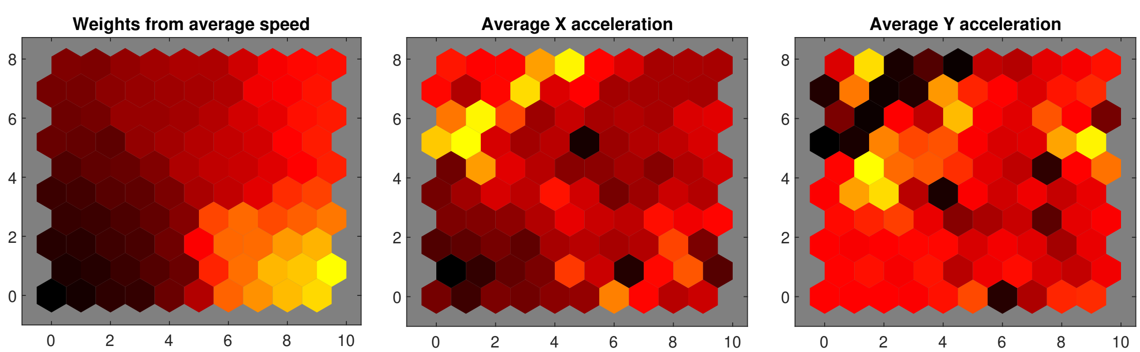
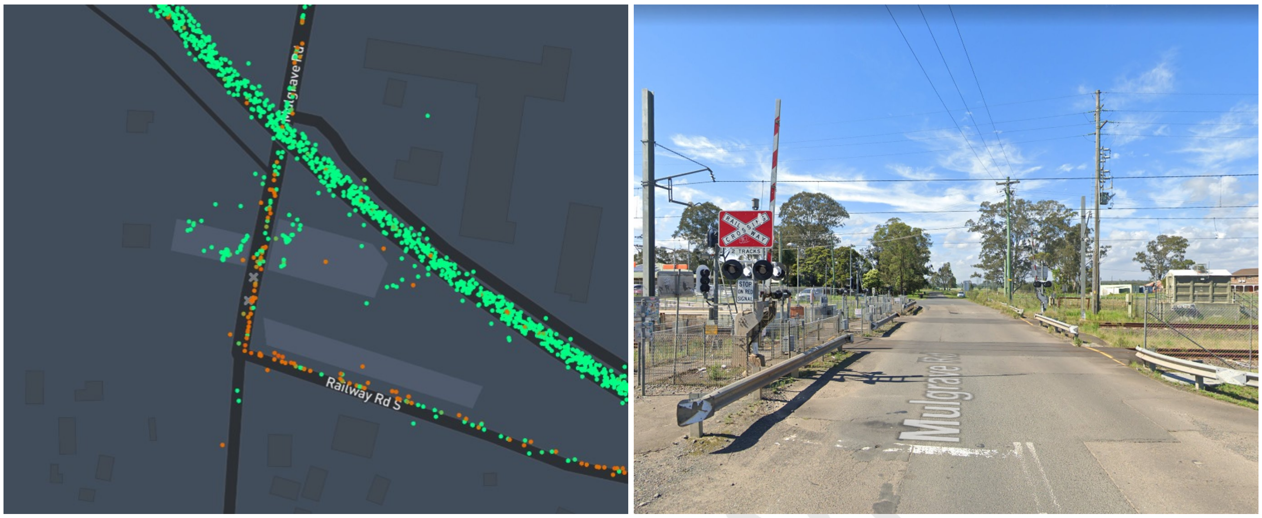


FIGURE 13: Weights in the neural network clustering.

To further investigate the reasons for the harsh driving behaviour, we look at the Google maps street view and the travel speed and traffic volumes from the Compass brake-point Data at the RLC on Mulgrave Road, Mulgrave, which includes 19 near misses from March 2020 to August 2021.

We discovered two significant manoeuvres entering the RLC and turning left after leaving the RLC. Both manoeuvres contain extreme deceleration in several cases of near misses. We described critical manoeuvres points and Google street view in Fig. 14.



(a) The recorded trajectory points of vehicles (b) Google street view

FIGURE 14: The selected direction at the RLC on Mulgrave Road

In Fig. 14a, green points represent safe driving behaviour, whereas red points denote less driving behaviour. Red points are primarily distributed in the vicinity of directions entering the RLC and turning left after leaving the RLC on Mulgrave Road. Moreover, we found that various X and Y acceleration points are greater than 0.47 and less than -0.47 among red points. This means that two significant manoeuvres entering the RLC and turning left after leaving the RLC entail sudden deceleration and acceleration manoeuvres in the vicinity of the RLC on Mulgrave Road.

Final risk index score

In this section, we produced a novel risk index score of RLCs in NSW. We normalised the number of historical incidents nearby RLCs, the service frequency of trains at RLCs, and the number of near misses in the vicinity of RLCs. We integrated the normalised values into one index score by varying each weight. The proposed formula is given as follows:

$$r_i = \alpha x_i + \beta y_i + \gamma z_i \tag{2}$$

where r_i, x_i, y_i, and z_i denote final risk index score, the normalised historical incidents, the normalised service frequency, and the normalised near misses, respectively, at ith RLC. α, β, and γ represent weight values corresponding to each normalised index. We calculated the mean and the standard deviation values for x_i, y_i, and z_i are 1.29, 5.88, 7.37, 19.02, and 0.13, 0.77, respectively, for a normalising process. We varied three weight values from 0.0 to 1.0 by 0.2, and the total of three weights is 1.0. We constructed 19 cases of combinations, including evenly distributed weights (i.e. all values are equal to 0.33). We identified the top 20 safety hotspots of RLCs from points of view, including historical incidents, service frequency, and near misses described in Table 4.

ID	ROAD_NAME	LGANAME	SUBURB	Incidents	Rank	Frequency	Rank	Near misses	Rank	x_i	y_i	z_i
945	Pine Road	CUMBERLAND	FAIRFIELD	85	2	241	1	7	3	1.00	1.00	1.00
53	Mulgrave Road	HAWKESBURY	MULGRAVE	21	14	72	10	19	1	1.00	1.00	1.00
350	Nolan Street	WOLLONGONG	UNANDERRA	49	6	67	17	2	15	1.00	1.00	0.99
644	Clarinda Street	BLUE MOUNTAINS	FAULCONBRIDGE	108	1	65	21	2	15	1.00	1.00	0.99
348	Bellambi Lane	WOLLONGONG	BELLAMBI	8	39	102	2	3	11	0.87	1.00	1.00
51	Bandon Road	BLACKTOWN	VINEYARD	8	39	72	10	5	6	0.87	1.00	1.00
645	Bundarra Street	BLUE MOUNTAINS	BLACKHEATH	50	5	58	26	1	31	1.00	1.00	0.87
56	Bourke Street	HAWKESBURY	RICHMOND	8	39	72	10	2	15	0.87	1.00	0.99
437	St James Road	NEWCASTLE	ADAMSTOWN	33	9	48	33	1	31	1.00	0.98	0.87
347	Park Road	WOLLONGONG	WOONONA	7	50	102	2	2	15	0.83	1.00	0.99
435	Glenrock Parade	CENTRAL COAST	KOOLEWONG	11	30	102	2	1	31	0.95	1.00	0.87
351	Princes Highway	WOLLONGONG	UNANDERRA	8	39	67	17	1	31	0.87	1.00	0.87
52	Level Crossing Road	HAWKESBURY	VINEYARD	3	115	72	10	2	15	0.61	1.00	0.99
1494	Darkes Road	WOLLONGONG	DAPTO	1	195	67	17	3	11	0.48	1.00	1.00
352	West Dapto Road	WOLLONGONG	KEMBLA GRANGE	1	195	67	17	2	15	0.48	1.00	0.99
434	Gosford Road / Rawson Road	CENTRAL COAST	WOY WOY	43	7	102	2	0	N/A	1.00	1.00	0.43
770	Beaumont Street	NEWCASTLE	HAMILTON	29	12	87	7	0	N/A	1.00	1.00	0.43
50	Garfield Road	BLACKTOWN	RIVERSTONE	76	3	72	10	0	N/A	1.00	1.00	0.43
55	Racecourse Road	HAWKESBURY	CLARENDON	38	8	72	10	0	N/A	1.00	1.00	0.43
637	Tynan Road	ALBURY CITY	TABLE TOP	24	13	4	103	3	11	1.00	0.43	1.00

TABLE 4: The top 20 safety hotspots according to the final risk index with even weights.

The RLC on Pine Road is the top-ranked due to frequent passenger train service, high historical incident rates, and the high chance of near misses in the vicinity of the RLC. Moreover, the RLC on Mulgrave Road is the second top-ranked due to its high chance of near misses, although the RLC is located in a regional area of NSW. Other RLCs are usually highly ranked in each criterion except for the RLCs on Gosford Road, Beaumont Street, Garfield Road, and Racecourse Road. Even if the number of near misses is zero nearby these four RLCs, the number of historical incidents and service frequency is relatively higher than other RLCs in NSW. The varied risk index scores combined with the weights provide the sensitivity of the final risk index scores of the vulnerable RLCs according to a case containing the different values of weights corresponding to each dynamic feature. We confirmed that the risk index scores of the top-ranked vulnerable RLCs are continuously sustained regardless of varied weights to each dynamic feature.

CONCLUSIONS

The ALCAM has been widely used in Australia and New Zealand to identify potential risks at level crossings and assist in prioritising crossing for upgrades in Australia and New Zealand due to its reliable and robust assessment mechanism. In this study, We newly introduced dynamic characteristics of motorists in the vicinity of RLCs and factors involved in the ALCAM. The dynamics include real-time train movements, the severity of traffic incidents nearby RLCs, and risky manoeuvres of motorists in the vicinity of RLCs. We identified hotspots of RLCs in NSW by integrating the three dynamic elements into the nominated weightings. The proposed study includes three contributions. First of all, investigating the microscopic dynamics of motorists and the severity of traffic incidents enables the introduction of near misses in the vicinity of RLCs in a safety assessment mechanism. Second, the data-driven analysis based on open-sourced data takes full advantage of data accessibility and flexibility if a large amount of data is secured. Last, the clustering analysis helps to understand the varied influence of each dynamic factor to define the risky behaviour of drivers nearby RLCs.

The proposed method lays the foundation stone for applying dynamic features of motorists to assess traffic safety of RLCs using public open data. It paves the way for analysing the historical behaviour of motorists in the vicinity of RLCs to anticipate key risks at each RLC in Australia. For future research directions, the proposed method will include temporal analysis of motorists' dynamics, including speed and XY accelerations, to assess the time-series evolution of dynamics according to implementing traffic safety facilities at RLCs in NSW. Besides, the applicability of the data-based approaches will be studied to construct flexible and robust safety assessment procedures at RLCs.

ACKNOWLEDGEMENT

The authors of this work would like to thank TfNSW for all the guidance and support provided. We make the observation that the current research index calculation are the team's own research contribution and findings, based on observed public data sets.

REFERENCES

1. Salmon, P. M., G. J. Read, N. A. Stanton, and M. G. Lenné, The crash at Kerang: Investigating systemic and psychological factors leading to unintentional non-compliance at rail level crossings. *Accident Analysis & Prevention*, Vol. 50, 2013, pp. 1278–1288.
2. Spicer, T. (ed.), *The Australian Level Crossing Assessment Model (ALCAM)*, AusRAIL PLUS 2007, Sydney, NSW, Australia, 2007.
3. Wigglesworth, E., Human factors in level crossing accidents. *Accident Analysis & Prevention*, Vol. 10, No. 3, 1978, pp. 229–240.
4. Tenkink, E. and R. Van der Horst, Car driver behavior at flashing light railroad grade crossings. *Accident Analysis & Prevention*, Vol. 22, No. 3, 1990, pp. 229–239.
5. Ghazel, M., Using Stochastic Petri Nets for Level-Crossing Collision Risk Assessment. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 10, No. 4, 2009, pp. 668–677.
6. Tey, L.-S., L. Ferreira, and A. Wallace, Measuring driver responses at railway level crossings. *Accident Analysis & Prevention*, Vol. 43, No. 6, 2011, pp. 2134–2141.
7. Mekki, A., M. Ghazel, and A. Toguyeni, Validation of a New Functional Design of Automatic Protection Systems at Level Crossings with Model-Checking Techniques. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 13, No. 2, 2012, pp. 714–723.
8. Salmon, P. M., M. G. Lenné, K. L. Young, and G. H. Walker, An on-road network analysis-based approach to studying driver situation awareness at rail level crossings. *Accident Analysis & Prevention*, Vol. 58, 2013, pp. 195–205.
9. Evans, A. W., Fatal accidents at railway level crossings in Great Britain 1946–2009. *Accident Analysis & Prevention*, Vol. 43, No. 5, 2011, pp. 1837–1845.
10. Evans, A. W. and P. Hughes, Traverses, delays and fatalities at railway level crossings in Great Britain. *Accident Analysis & Prevention*, Vol. 129, 2019, pp. 66–75.
11. Zhao, S. and A. Khattak, Motor vehicle drivers' injuries in train–motor vehicle crashes. *Accident Analysis & Prevention*, Vol. 74, 2015, pp. 162–168.
12. Liang, C., M. Ghazel, O. Cazier, and E.-M. El-Koursi, A new insight on the risky behavior of motorists at railway level crossings: An observational field study. *Accident Analysis & Prevention*, Vol. 108, 2017, pp. 181–188.
13. Gitelman, V. and A. Hakkert, The evaluation of road-rail crossing safety with limited accident statistics. *Accident Analysis & Prevention*, Vol. 29, No. 2, 1997, pp. 171–179.
14. Russell, E. R. and M. I. Mutabazi, Rail-Highway Grade Crossing Consolidation in Kansas. *Transportation Research Record*, Vol. 1648, No. 1, 1998, pp. 1–7.
15. Miranda-Moreno, L. F., L. Fu, F. F. Saccomanno, and A. Labbe, Alternative Risk Models for Ranking Locations for Safety Improvement. *Transportation Research Record*, Vol. 1908, No. 1, 2005, pp. 1–8.
16. Hu, S.-R. and K.-H. Wu, Accident Risk Analysis and Model Applications of Railway Level Crossings. In *2008 11th International IEEE Conference on Intelligent Transportation Systems*, 2008, pp. 687–692.
17. Gruyter, C. D. and G. Currie, Rail-road crossing impacts: an international synthesis. *Transport Reviews*, Vol. 36, No. 6, 2016, pp. 793–815.
18. Soleimani, S., S. R. Mousa, J. Codjoe, and M. Leitner, A Comprehensive Railroad-Highway Grade Crossing Consolidation Model: A Machine Learning Approach. *Accident Analysis & Prevention*, Vol. 128, 2019, pp. 65–77.

19. Haleem, K., Investigating risk factors of traffic casualties at private highway-railroad grade crossings in the United States. *Accident Analysis & Prevention*, Vol. 95, 2016, pp. 274–283.

20. Keramati, A., P. Lu, A. Iranitalab, D. Pan, and Y. Huang, A crash severity analysis at highway-rail grade crossings: The random survival forest method. *Accident Analysis & Prevention*, Vol. 144, 2020, p. 105683.

21. Keramati, A., P. Lu, D. Tolliver, and X. Wang, Geometric effect analysis of highway-rail grade crossing safety performance. *Accident Analysis & Prevention*, Vol. 138, 2020, p. 105470.

22. Lee, S., D. Ngoduy, and M. Keyvan-Ekbatani, Integrated deep learning and stochastic car-following model for traffic dynamics on multi-lane freeways. *Transportation research part C: emerging technologies*, Vol. 106, 2019, pp. 360–377.

23. Lee, S., S. C. Wong, C. C. C. Pang, and K. Choi, Real-Time Estimation of Lane-to-Lane Turning Flows at Isolated Signalized Junctions. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 16, No. 3, 2015, pp. 1549–1558.

24. Lee, S., S. Wong, and Y. Li, Real-time estimation of lane-based queue lengths at isolated signalized junctions. *Transportation Research Part C: Emerging Technologies*, Vol. 56, 2015, pp. 1–17.

25. Ngoduy, D., S. Lee, M. Treiber, M. Keyvan-Ekbatani, and H. Vu, Langevin method for a continuous stochastic car-following model and its stability conditions. *Transportation Research Part C: Emerging Technologies*, Vol. 105, 2019, pp. 599–610.