# TRB Annual Meeting

## CONSIDERING MICROSCOPIC BEHAVIOUR OF MOTORISTS IN SAFETY ASSESSMENT FOR RAILWAY LEVEL CROSSING IN AUSTRALIA
--Manuscript Draft--

| | |
|---|---|
| **Full Title:** | CONSIDERING MICROSCOPIC BEHAVIOUR OF MOTORISTS IN SAFETY ASSESSMENT FOR RAILWAY LEVEL CROSSING IN AUSTRALIA |
| **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. |
| **Manuscript Classifications:** | Rail; Highway/Rail Grade Crossings AR080; General; Rail Safety AR070; Human Factors in Crashes; Safety; Safety Performance and Analysis ACS20; Crash Analysis; Transportation Safety Management Systems ACS10; Behavioral Safety Analysis and Program Development |
| **Manuscript Number:** | TRBAM-23-01377 |
| **Article Type:** | Presentation |
| **Order of Authors:** | Seunghyeon Lee, PhD |
| | Tuo Mao, PhD |
| | Yuming Ou, PhD |
| | Adriana-Simona Mihaita, PhD |
| | Fang Chen, PhD |
| **Additional Information:** | |
| **Question** | **Response** |
| The total word count limit is 7500 words including tables. Each table equals 250 words and must be included in your count. Papers exceeding the word limit may be rejected. My word count is: | 7488 |
| Is your submission in response to a Call for Papers? (This is not required and will not affect your likelihood of publication.) | Yes |
| Please specify which Call for Papers by selecting from the drop down list. View the more information about the Call for Papers here. as follow-up to "Is your submission in | AP080 Call for Papers - AP080, Standing Committee on Transit Safety and Security |

| response to a Call for Papers? (This is not required and will not affect your likelihood of publication.)" | |

1 **CONSIDERING MICROSCOPIC BEHAVIOUR OF MOTORISTS IN SAFETY**
2 **ASSESSMENT FOR RAILWAY LEVEL CROSSING IN AUSTRALIA**
3
4
5
6 **Seunghyeon Lee, Ph.D., Corresponding Author**
7 Assistant Professor
8 Department of Transportation Engineering
9 University of Seoul, Seoul, Korea, 02504
10 Email: seunghyeon.lee@uos.ac.kr
11
12 **Tuo Mao, Ph.D.**
13 Faculty of Engineering and IT
14 University of Technology Sydney, Ultimo NSW 2007, Australia
15 Email: tuo.mao@uts.edu.au
16
17 **Yuming Ou, Ph.D.**
18 Faculty of Engineering and IT
19 University of Technology Sydney, Ultimo NSW 2007, Australia
20 Email: yuming.ou@uts.edu.au
21
22 **Adriana-Simona Mihaita, Ph.D.**
23 Faculty of Engineering and IT
24 University of Technology Sydney, Ultimo NSW 2007, Australia
25 Email: adriana-simona.mihaita@uts.edu.au
26
27 **Fang Chen, Ph.D.**
28 Faculty of Engineering and IT
29 University of Technology Sydney, Ultimo NSW 2007, Australia
30 Email: fang.chen@uts.edu.au
31
32
33
34
35 Word Count: 6488 words + 4 table(s) × 250 = 7488 words
36
37
38
39
40
41
42 Submission Date: July 30, 2022

1 **ABSTRACT**

2 This work proposes a methodology considering the dynamic features of motorists in the vicinity
3 of railway level crossings (RLCs) to the Australian Level Crossing Assessment Model (ALCAM).
4 The dynamic features include real-time train movements, the severity of traffic incidents nearby
5 RLCs, and the risky manoeuvres of motorists when crossing RLCs. We collected train timeta-
6 bles and network data provided by the TfNSW Open Data Hub and real-time trajectories of trains
7 provided by GTFS feeds (General Transit Feed specification) to demonstrate real-time train move-
8 ments around level crossings. To quantify the severity of traffic incidents in the RLC vicinity, we
9 used NSW road crash data sets, including crash, number of victims, and unit profiles. To identify
10 driving risky manoeuvres, we investigated the speed, XY acceleration, and the G-force of road ve-
11 hicles to define safe-points and brake-points in the vicinity of RLCs. Furthermore, we employed a
12 neural network clustering method to identify dynamic characteristics of each safe-point and brake-
13 point in the vicinity of RLCs. We integrated three dynamic elements to the nominated weightings
14 with sensitivity analysis. The proposed method lays the foundation stone for considering dynamic
15 features of motorists and assess the rail crossing safety by using public open data. It paves the way
16 for analysing the historical behaviour of motorists in the vicinity of RLCs and anticipate key risks
17 at each RLC in Australia.
18
19 *Keywords*: Safety assessment, Railway level crossing, Behaviour of motorists, Neural network
20 clustering

1 **INTRODUCTION**
2 In Australia, securing safety at railway level crossings (RLCs), where motorists and trains intersect
3 at the same grade, is one of the crucial concerns in an integrated multi-modal transport network.
4 Crashes at RLCs have involved severe fatalities, injuries, and came with significant economic ex-
5 penses even if the number of crashes at the RLC constitutes below 1% of total crashes in New South
6 Wales (NSW) in Australia for the last several decades. For example, Salmon et al. (*1*) analysed the
7 systemic and psychological factors causing the tragedy at an RLC near Kerang, Victoria, Australia,
8 on 5 June 2007. Eleven train passengers were killed by a loaded semi-trailer truck striking a pas-
9 senger train at the RLC. This incident additionally involved 13 injured passengers and an injured
10 truck driver. They used two juxtaposed human factors approaches to produce insights into the
11 contributory factors underlying the incident. They concluded that the tragic event could be caused
12 by the lack of fully active controls from the systemic perspective and a Looked-But-Failed-To-See
13 (LBFS) error triggered by an inappropriate schema from the psychological perspective.
14    To systematically assess the vulnerability of an individual RLC location preventing prevent
15 recurrence of the tragedy in Australia, the national ALCAM group developed the Australian Level
16 Crossing Assessment Model (ALCAM), which is an assessment tool used to identify key potential
17 risks at level crossings and to assist in the prioritisation of crossings for upgrades in (*2*). This
18 complex scoring process is outlined in Australian Standard and New Zealand Standard (AS/NZS
19 4360:2004) includes a matrix of weightings that measures the influence of the designated char-
20 acteristics at RLCs on the possible accident mechanisms. ALCAM continues to be updated with
21 fine-tuning weightings to consider new control technology and modifications. Due to its reliable
22 and robust assessment mechanism for RLCs, the Australian Transport Council of Commonwealth,
23 State and Territory, the Standing Committee of Transport (SCOT), all state and territory transport
24 ministers, and New Zealand agreed to adopt the ALCAM to identify potential accident causal
25 factors and overall effects of proposed treatments.
26    This study aims to introduce the dynamic characteristics of motorists and trains in addition
27 to the factors already involved in the ALCAM. We used a data-driven analysis based on NSW
28 open-sourced data to identify hot spots of RLCs in NSW, Australia. In this study, the dynamic
29 characteristics include real-time train movements, the severity of traffic incidents nearby RLCs,
30 and risky manoeuvres of motorists nearby RLCs. We collected train timetables and network data
31 from TfNSW Open Data Hub to explain the real-time train movements at an individual RLC lo-
32 cation. We used NSW road crash data sets, including crash, persons, and unit profiles, to estimate
33 the severity of traffic incidents nearby each RLC. We investigated the acceleration, deceleration,
34 G-force, and speed of road vehicles to understand risky manoeuvres of motorists defined by safe-
35 points and brake-points in the vicinity of RLCs. We identified hot spots of RLCs in NSW in the
36 proposed assessment method by integrating the three dynamic elements to the nominated weight-
37 ings with a sensitivity analysis. The specific contributions of the study are given below:
38    • investigating the microscopic dynamics of motorists and the severity of traffic incidents
39      enables to introduce **near misses** in the vicinity of RLCs in a safety assessment mecha-
40      nism.
41    • the data-driven analysis based on open-sourced data takes full advantage of **data acces-**
42      **sibility and flexibility** if a large amount of data is secured.
43    • **the sensitivity analysis** helps to understand the varied influence of each dynamic factor
44      on the identification of hotspots of RLCs.
45    To achieve the research goal, this article is organised as follows. The literature review for

1 safety assessment methods for RLCs and data profiles is presented in Sections 2 and 3, respec-
2 tively. The assessment methods and results for dynamics of motorists and trains nearby RLCs are
3 demonstrated in Section 4. Finally, Section 6 provides the conclusions of this study and future
4 research directions.

5 **LITERATURE REVIEW**
6 This section summarises the studies on railway level crossings, considering the behaviour of mo-
7 torists and trains in an RLC safety assessment and identifies the research gaps this study tackled.

8 **Behaviour of motorists in the vicinity of RLCs**
9 RLCs have been regarded as hotspots of reduced traffic safety in multi-modal urban networks
10 since a single crash between trains and vehicles has brought about its swingeing damage at RLCs.
11 Understanding behaviour of motorists in the vicinity of RLCs is one of the key factors to secure
12 safe operating and driving environment at RLCs in (*3*), (*4*), (*5*), (*6*), (*7*), and (*8*).
13         Tenkink and Van der Horst (*4*) examined the behaviour of car drivers at two RLCs, in
14 which flashing warning lights are installed using analysing recorded video. Evans (*9*) investigated
15 fatal accidents and fatalities at RLCs in Great Britain from 1946 to 2009 from a macroscopic
16 perspective. They found that the number of fatalities per year have decreased by 65% in the first
17 half of the study period, whereas after that they have been relatively constant by the end of the
18 study period. Evans and Hughes (*10*) continued to conduct the safety investigation at RLCs from
19 a microscopic perspective in Great Britain. They investigated relationships between traverses,
20 delays, and fatalities to road users at RLCs to confirm an additional risk for the users.
21         Salmon et al. (*8*) mentioned that the situation awareness and the decision-making of mo-
22 torists are not sufficiently considered in existing methodologies to evaluate the safety performance
23 of RLCs. They conducted an on-road network analysis to examine the situation awareness of mo-
24 torists at RLCs. Zhao and Khattak (*11*) investigated injuries of motorists involved in train-motor
25 vehicle crashes at RLCs using three injury severity models based on ordered probit, multinomial
26 logit, and random parameter logit. They found that the likelihood of more severe crashes increases
27 when they involve higher train and vehicle speeds, freight trains, older drivers, and female drivers.
28 Liang et al. (*12*) analysed the macroscopic behaviour of motorists while crossing RLCs during the
29 closure cycle. They used the violation rate of speeding to define the risky behaviour of vehicle
30 drivers estimated during RLCs closure cycles.

31 **Safety assessment for RLCs**
32 Gitelman and Hakkert (*13*) used a hazard index to evaluate the RLCs safety based on limited histor-
33 ical accident statistics in Israel. Russell and Mutabazi (*14*) adopted to calculate a designated rating
34 for each considered feature to construct a ranking table for RLCs closures in Kansas, the United
35 States. Miranda-Moreno et al. (*15*) evaluated Canadian RLCs based on an accident history of 5
36 years using the traditional negative binomial model, the heterogeneous negative binomial model,
37 and the Poisson log-normal model. They confirmed that a list of hazardous RLCs significantly de-
38 pends on the model assumptions and raking criteria. Hu and Wu (*16*) conducted an empirical study
39 to identify accident frequency and severity. They employed a logit model to categorise severity into
40 three levels: no-severity, minor-severity, and serious-severity. Gruyter and Currie (*17*) carried out a
41 detailed literature review to develop an international synthesis of RLCs impacts. They revealed 18
42 different types of impacts associated with RLCs; meanwhile, safety defined by accident frequency

1 and severity at RLCs is one of the essential factors in an RLC safety assessment process. They as-
2 serted that using empirical evidence to support impact assessments is required to understand RLCs
3 impacts better.

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

## Research gaps

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

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

34       Second, open-sourced data sets were not attractive options in safety assessment methods
35 of RLCs due to data quality, reliability, and accessibility. In recent years, the debut of informa-
36 tion and communication technologies (ICT) in traffic engineering has diversified spatial-temporal
37 dimensions and resolutions of available data sets for transport systems analysis. A multitude of
38 traffic data sources is available on an urban transportation network to improve temporal and spatial
39 coverages of intermodal traffic patterns in real-time. The multiple data-sources include fixed sen-
40 sors and mobility sensors in (23), (24), (22) and (25). Conventional fixed sensors are a relatively
41 reliable and robust-data source, whereas expensive construction and maintenance costs are inher-
42 ent in this type of sensor. In the meantime, personal mobility data sources can provide the detailed
43 spatial and temporal behaviour of moving objects' movements, although this type of data requires
44 effective statistical sampling techniques because of the sparsity of available data. In this study, we

1  used train timetables and crash data sets obtained from TfNSW Open Data Hub and FCD derived
2  from GPS data from Compass IoT in NSW to maximise data accessibility. Moreover, we con-
3  ducted the sensitivity analysis among safety features at RLCs to understand the varied influence of
4  each dynamic factor on the identification of hotspots of RLCs and maximise the flexibility of data
5  analysis.

6  **DATA ANALYSIS**
7  We secured five data sets from four different data sources. The first data is an information sheet
8  about RLCs obtained from NSW public level crossing finder[1]. The information sheet contains
9  the profiles of 1360 RLCs, including identification number, road name, control type, owner, line
10 section, suburb, number of tracks, longitude, and latitude.
11     The second data is a historical unplanned incident profile from 16 August 2006 to 11 Au-
12 gust 2021 obtained from Traffic Management Centre (TMC) in NSW. The historical incident pro-
13 file involves over a million incident records that have been recorded together with information
14 regarding location, duration, type, link ID, segment ID, direction, and Sector ID. These historical
15 incident records allowed us to access high-resolution location information of a single incident for
16 hot-spots identification of RLCs; meanwhile, the severity of crashes was not provided.
17     The third data is a historical crash profile from 1 January 2015 to 31 December 2019 pro-
18 vided by TfNSW Open Data Hub[2]. This data set contains crash profiles and information on the
19 people and units involved. In crash profiles, they provided a degree of the crash, street of the crash,
20 type of location, primary permanent, temporary, and hazardous features, surface conditions, natu-
21 ral lights, signals operations, speed limit, users movement, short description, first impact type, the
22 number of traffic units involved, and the number of killed, the number of serious, moderately, and
23 minor-other injured. They linked crash profiles to profiles of people involved, including gender,
24 age group, road user class, degree of casualty, and units, containing a role in the first impact, types,
25 direction of travel, manoeuvre and the first and the second crashed objects.
26     The fourth data consists of railway timetables and network profiles from 1 January 2019
27 to 31 August 2021 obtained from TfNSW Open Data Hub and Google GTFS. Railway timetables
28 provide information on a train service schedule, including departure station, departure time, arrival
29 station and arrival time for each passenger train service. The Railway network includes information
30 on train line geometry and train station location.
31     The fifth data is microscopic dynamics of motorists from 1 March 2021 to 31 August
32 2021 provided by Compass IoT[3]. Compass IoT is a data aggregation and analytic company that
33 uses connected vehicles to generate in-depth insights into vehicular trajectories across transport
34 networks. They collected trajectories data from private and public data providers to provide the
35 average speed profiles and volumes in a single link across the Australian road network. The col-
36 lected microscopic dynamics of motorists enable us to evaluate the driving behaviour of vehicles
37 in the vicinity of RLCs by analysing vehicular manoeuvres, including braking, acceleration, and
38 deceleration of a single-vehicle.
39     These five data sets were used to analyse profiles of RLCs, traffic incidents, railway opera-
40 tions, 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/

1 **Railway level crossing**
2 We investigated general information on RLCs in NSW. The data set of train level crossings in
3 NSW contains the following features: CROSSING NUMBER, ROAD NAME, LOCATION, CON-
4 TROL TYPE, CONTROL CATEGORY, RAIL OWNER, LINE SECTION, LGA NAME, SUB-
5 URB, RAIL KM, NO OF TRACKS, OBJECT ID, lon, and lat. The features of lon and lat were
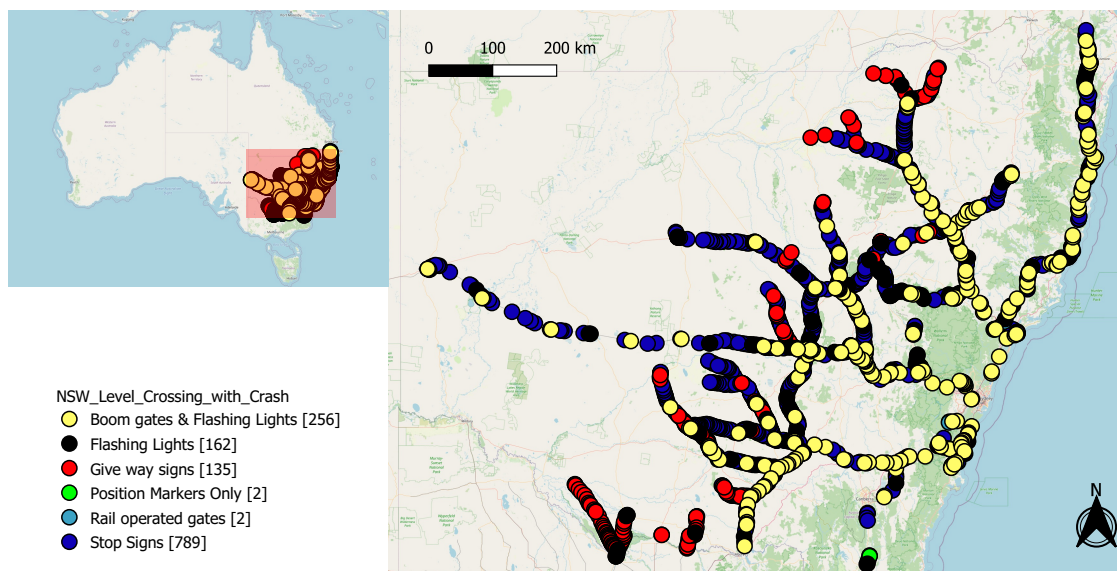6 used for map-matching in a QGIS program.



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

7       Active and passive types of level crossing controls are installed across NSW. In the active
8 type, boom gates and flashing lights are installed at 256 RLCs. Flashing lights are operated at
9 162 of RLCs across NSW. Rail-operated gates are installed at two places of RLCs. In the passive
10 type, stop signs and give way signs are installed at 789 and 135 RLCs, respectively. Fig. 1 presents
11 distributed locations of each level crossing control category at RLCs. Stop signs in the passive type
12 are widely installed across regional NSW, whereas boom gates and flashing lights in the active type
13 are mainly installed in the Sydney metropolitan area.

14 **Traffic incidents**
15 We investigated two open-sourced traffic incident data sets, including historical unplanned inci-
16 dent profiles over ten years and crash profiles for five years. Unplanned incident logs present the
17 summarised information of over a million incidents in NSW, whereas crash profiles illustrate the
18 detailed information of each crash of over 100 thousand. In unplanned incident logs, we filtered
19 this data set to extract only train-related incidents, and the remaining findings revealed that 2,549
20 were train-related incidents. The data revealed an average duration of 139 minutes or train dis-
21 ruption, representing quite a considerable time interval with no train service in affected areas or
22 trains blocked in specific train networks without being able to continue their journey. The maxi-
23 mum duration of a train incident has reached 7,663 minutes and was signalled in Glenyalla north
24 of Kankool on 15 February 2015.
25       In crash profile data, we analysed crashes, units involved in crashes, and people involved
26 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
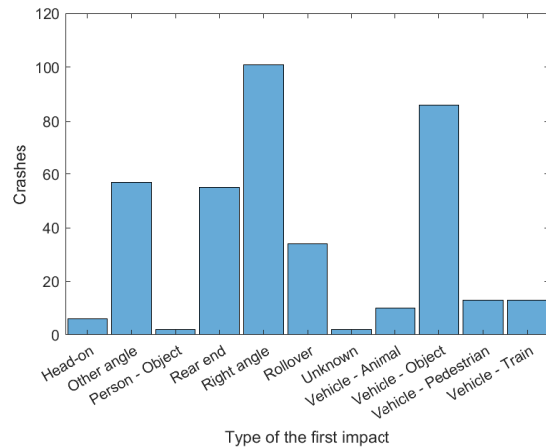crashes nearby RLCs in NSW (Fig. 2d).
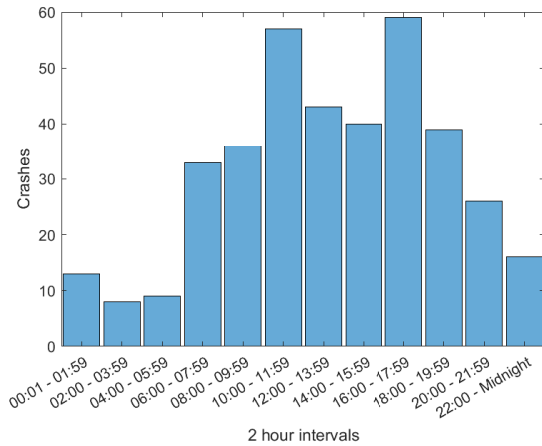
(a) 2 hours intervals
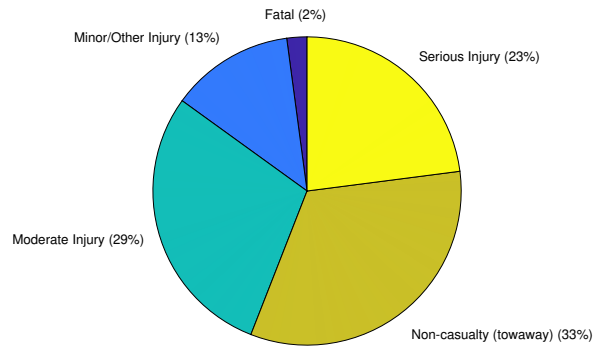


(b) Degree of crash
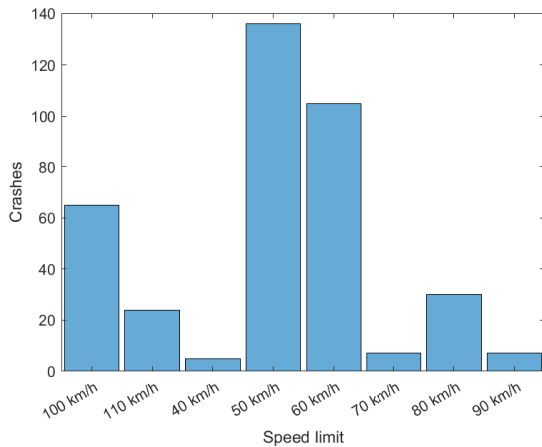


(c) Speed limit



(d) Type of the first crash

**FIGURE 2**: Crashes in the vicinity of RLCs

## 1 Railway operations

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

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

13     Three data sets were processed through two steps: map matching and arrival time estimat-
14 ing. In the map matching, the locations of the RLCs are not precisely located on the train lines
15 since the RLCs data and the train network data set are collected from different organisations. As

1   a result, we employed a map matching technique to find the correct location of the RLCs on the
2   train lines.
3           We identified 590 (43%) of RLCs, which are operating with the given timetables in the
4   passenger train line, while 770 (57%) of RLCs are not. They are illustrated in Fig. 3a and Fig. 3b,
5   respectively. In the estimating step of train arrival time at RLCs, we excluded 770 of RLCs from
6   the investigation because the given passenger timetable data do not cover them.
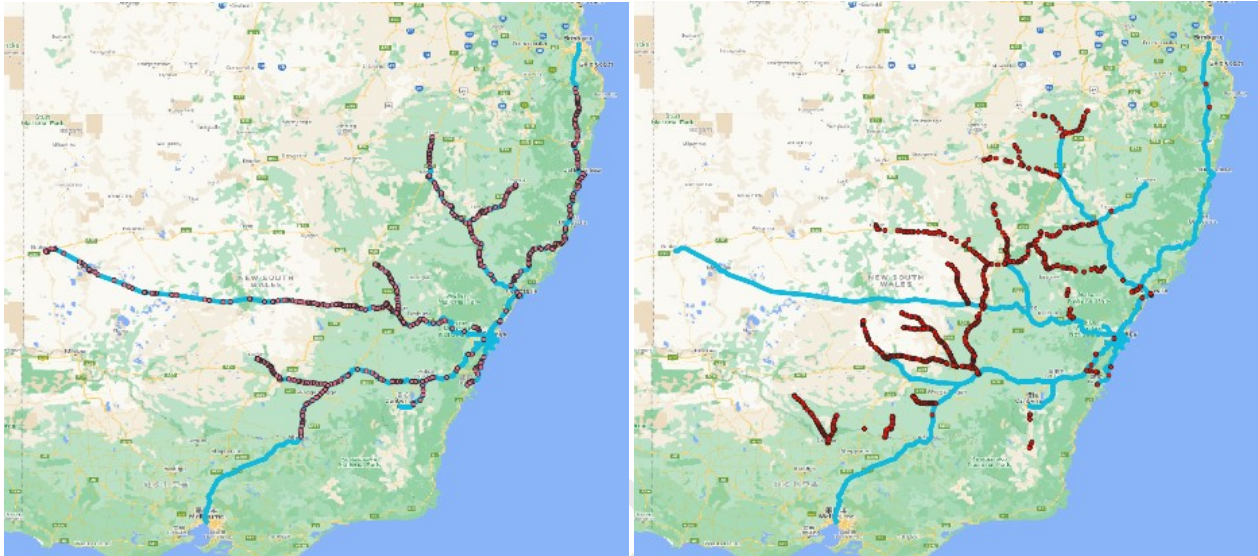


(a) RLCs that can be projected                              (b) RLCs that cannot be projected

**FIGURE 3**: The results of the map matching

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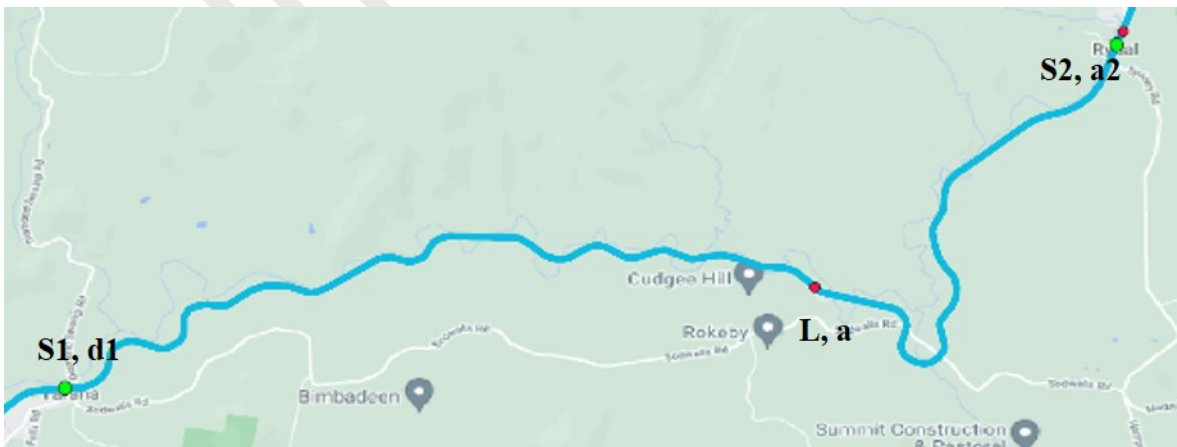


**FIGURE 4**: Estimating arrival time at RLCs.

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

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

1  where $d_1$ is the departure time at station $S_1$,

2  $a_2$ is the arrival time at station $S_1$,

3  $Distance_{S_1,L}$ is the distance from station $S_1$ to RLC $L$ along the train line,

4  and $Distance_{S_1,S_2}$ is the distance from station $S_1$ to station $S_2$ along the train line.

5  **Behaviour of motorists**

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

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

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

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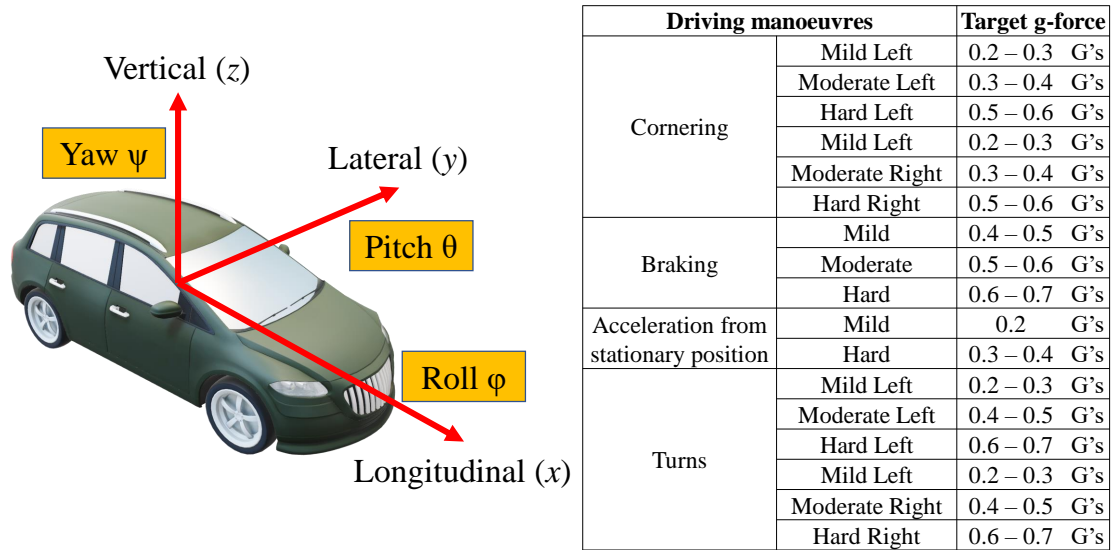33  dynamic data measurements.

| 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
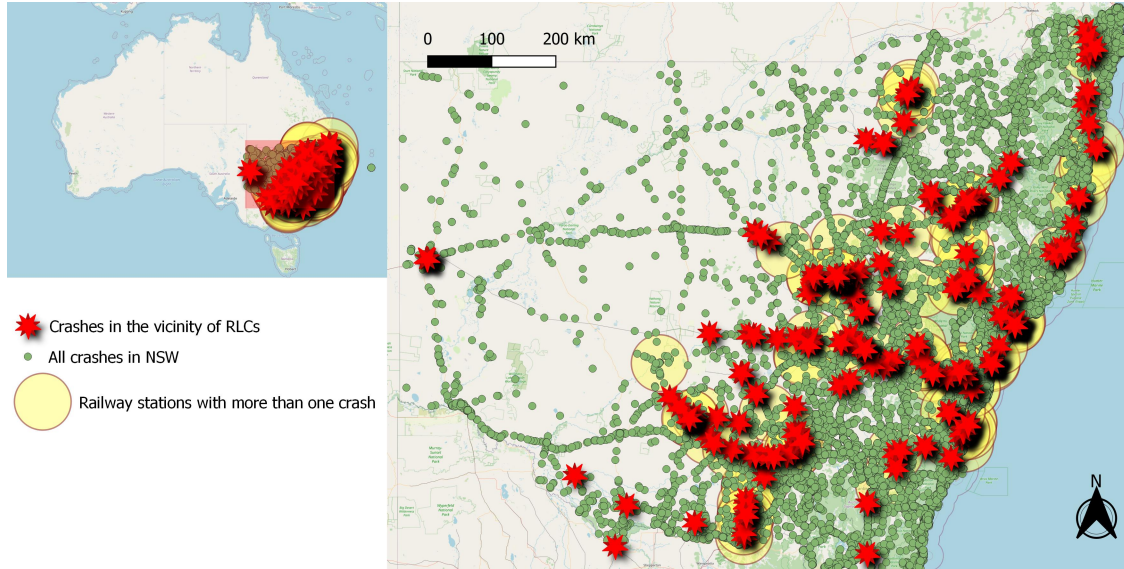units involved. The top ten safety hotspots are illustrated in Table 1.



**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.

1 **Railway operational factors**
2 We constructed the ranking tables for entire stations in NSW through railway operational data
3 analysis. In NSW, we observed 57 level crossings with a daily service frequency higher than ten
4 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

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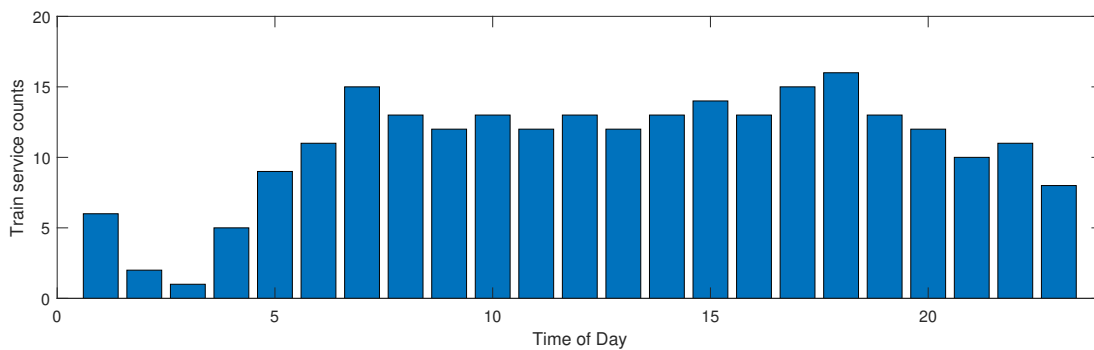


**FIGURE 7**: Hourly train service frequency at the RLC on Pine Road.

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

**Behaviour factors of motorists**

8   We present the top ten RLCs according to the number of incidents and near misses in a 150-metre
9   radius. In Table 3, all RLCs are controlled by an active type of control. Boom gates and flashing
10  lights control 9 out of 10 RLCs. We detect 18 near misses in the 150-metre vicinity of the RLC
11  on Mulgrave Road, Mulgrave. Furthermore, eight near misses were observed in the vicinity of the
12  RLC on Borenore Road and Amaroo Road, Borenore. Six RLCs involve over five near misses in
13  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.

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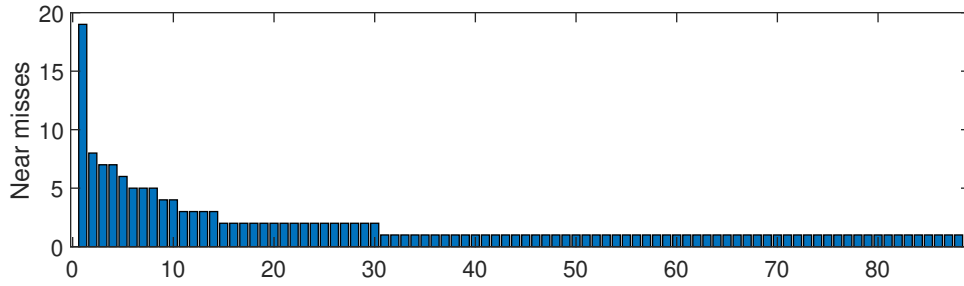17  Y acceleration are illustrated inFig. 9, Fig. 10, and Fig. 11.
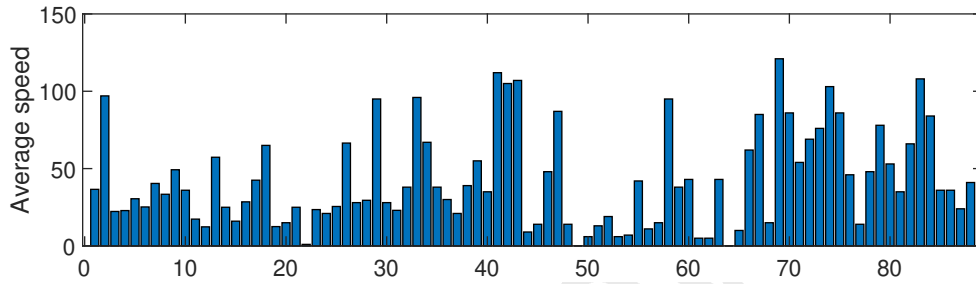
**FIGURE 8**: Near misses in a 150m radius.



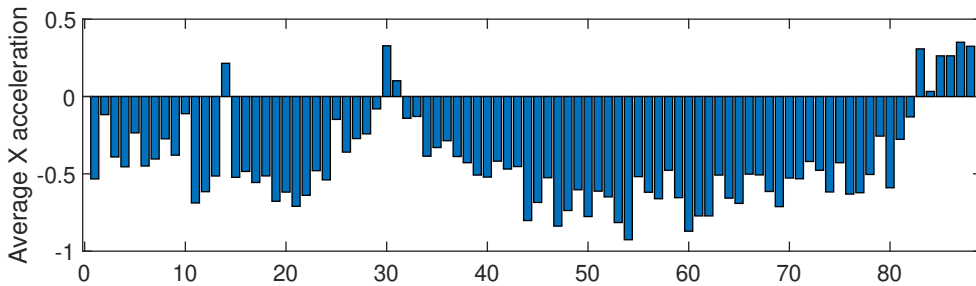**FIGURE 9**: Average speed.



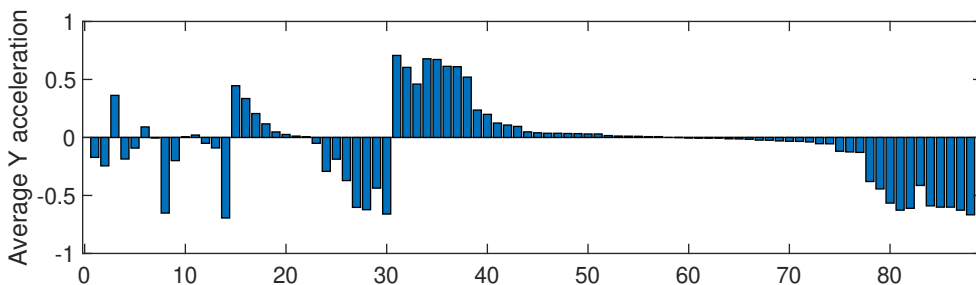**FIGURE 10**: Average X acceleration.



**FIGURE 11**: Average Y acceleration.

1        Based on the three dynamic features, we classified RLCs into five groups according to
2   the severity of the g-force. The severity of the g-force is represented in the platform as harsh

1 acceleration and harsh braking in Fig. 5. Deceleration or braking from 60 $km/h$ to 0 $km/h$ in 3.0
2 seconds at 0.6 $g$ is considered as harsh braking. Accelerating from 0 $km/h$ to 60 $km/h$ over 0.5 $g$'s
3 is hard acceleration. Swerving or cornering is considered harsh when it is over 0.47 $g$'s. The RLCs
4 in NSW are categorised into 18 RLCs as hard brake and stiff steering, 54 RLCs as hard brake and
5 no steering, 4 RLCs as lightweight brake and stiff steering, and 3 RLCs as lightweight brake and
6 no steering, and 9 RLCs as accelerating and stiff steering.

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

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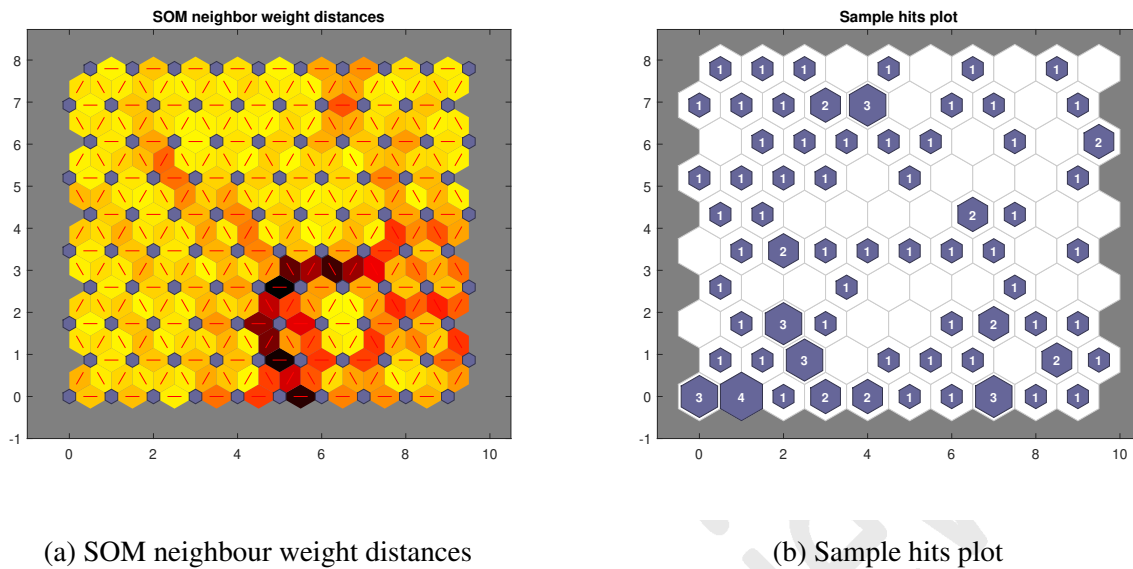
(a) SOM neighbour weight distances          (b) Sample hits plot

**FIGURE 12**: Self-organising maps with hits

1        Moreover, we demonstrated how many RLCs are associated with each of the neurons in
2  Fig. 12b. The maximum number of RLCs associated with a neuron is four, in which the neuron is
3  closely related to the surrounding neurons.
4        Fig. 13 describes a weight plane for average speed, average X acceleration, and average
5  Y acceleration at each RLC. These plots show the weight connecting each attribute to each neu-
6  ron, with darker colours showing larger weights. According to these plots, three features are not
7  correlated with each other; meanwhile, features play a significant role in clustering RLCs at each
8  platoon of neurons in this analysis. Furthermore, the clustering result by a neural network pro-
9  duced similar grouping results derived from analysing behaviour of incidents around RLCs based
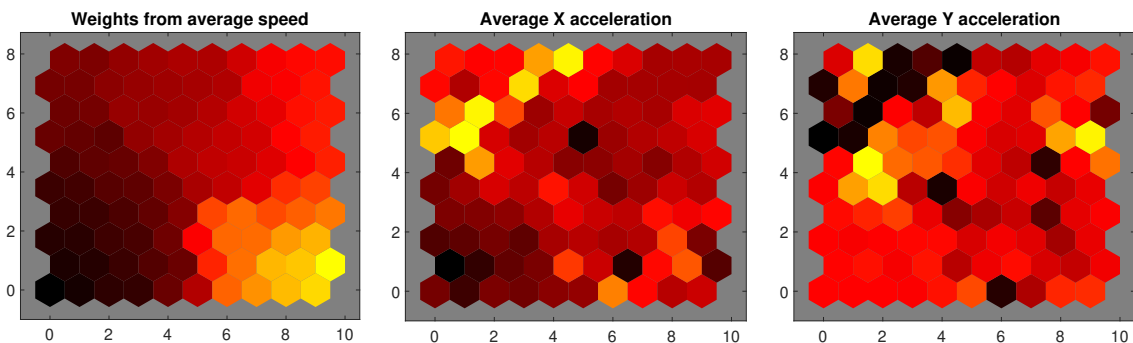10  on G-force.



**FIGURE 13**: Weights in the neural network clustering.

11        To further investigate the reasons for the harsh driving behaviour, we look at the Google
12  maps street view and the travel speed and traffic volumes from the Compass brake-point Data at
13  the RLC on Mulgrave Road, Mulgrave, which includes 19 near misses from March 2020 to August
14  2021.

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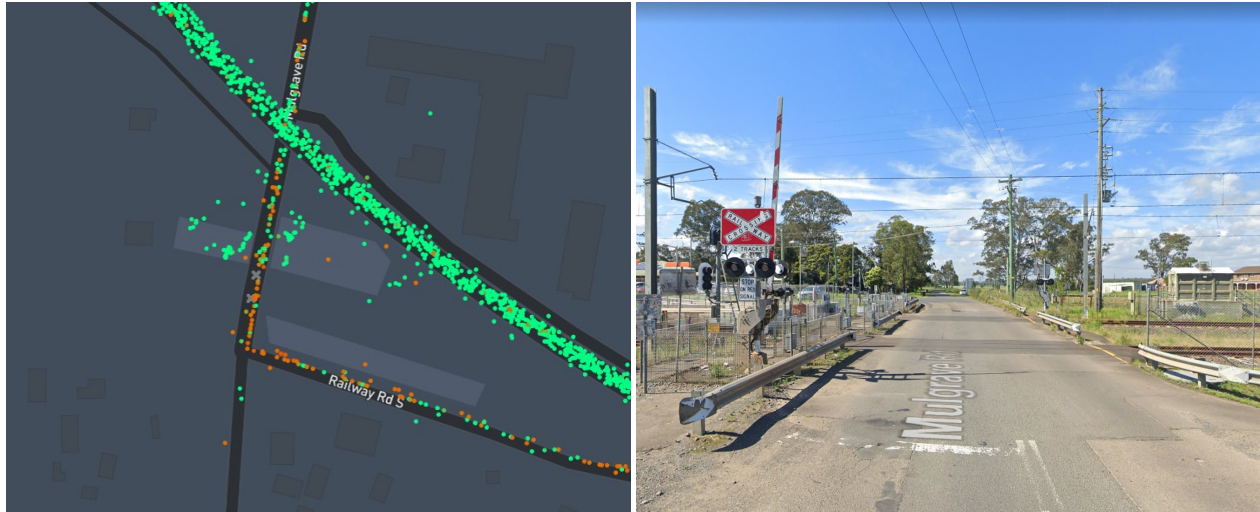


(a) The recorded trajectory points of vehicles                      (b) Google street view

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

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

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

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

15      where $r_i$, $x_i$, $y_i$, and $z_i$ denote final risk index score, the normalised historical incidents, the
16 normalised service frequency, and the normalised near misses, respectively, at $i$th RLC. $\alpha$, $\beta$, and
17 $\gamma$ represent weight values corresponding to each normalised index. We calculated the mean and the
18 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,
19 for a normalising process. We varied three weight values from 0.0 to 1.0 by 0.2, and the total of
20 three weights is 1.0. We constructed 19 cases of combinations, including evenly distributed weights
21 (i.e. all values are equal to 0.33). We identified the top 20 safety hotspots of RLCs from points of
22 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 | 0.99 |
| 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.

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

12  **CONCLUSIONS**
13  The ALCAM has been widely used in Australia and New Zealand to identify potential risks at
14  level crossings and assist in prioritising crossing for upgrades in Australia and New Zealand due
15  to its reliable and robust assessment mechanism. In this study, We newly introduced dynamic
16  characteristics of motorists in the vicinity of RLCs and factors involved in the ALCAM. The dy-
17  namics include real-time train movements, the severity of traffic incidents nearby RLCs, and risky
18  manoeuvres of motorists in the vicinity of RLCs. We identified hotspots of RLCs in NSW by
19  integrating the three dynamic elements into the nominated weightings. The proposed study in-
20  cludes three contributions. First of all, investigating the microscopic dynamics of motorists and
21  the severity of traffic incidents enables the introduction of near misses in the vicinity of RLCs in a
22  safety assessment mechanism. Second, the data-driven analysis based on open-sourced data takes
23  full advantage of data accessibility and flexibility if a large amount of data is secured. Last, the
24  clustering analysis helps to understand the varied influence of each dynamic factor to define the
25  risky behaviour of drivers nearby RLCs.
26    The proposed method lays the foundation stone for applying dynamic features of motorists
27  to assess traffic safety of RLCs using public open data. It paves the way for analysing the historical
28  behaviour of motorists in the vicinity of RLCs to anticipate key risks at each RLC in Australia.
29  For future research directions, the proposed method will include temporal analysis of motorists'

1 dynamics, including speed and XY accelerations, to assess the time-series evolution of dynamics
2 according to implementing traffic safety facilities at RLCs in NSW. Besides, the applicability of the
3 data-based approaches will be studied to construct flexible and robust safety assessment procedures
4 at RLCs.

5 **ACKNOWLEDGEMENT**

# 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.

1  19.   Haleem, K., Investigating risk factors of traffic casualties at private highway-railroad grade
2        crossings in the United States. *Accident Analysis & Prevention*, Vol. 95, 2016, pp. 274–
3        283.
4  20.   Keramati, A., P. Lu, A. Iranitalab, D. Pan, and Y. Huang, A crash severity analysis at
5        highway-rail grade crossings: The random survival forest method. *Accident Analysis &*
6        *Prevention*, Vol. 144, 2020, p. 105683.
7  21.   Keramati, A., P. Lu, D. Tolliver, and X. Wang, Geometric effect analysis of highway-rail
8        grade crossing safety performance. *Accident Analysis & Prevention*, Vol. 138, 2020, p.
9        105470.
10 22.   Lee, S., D. Ngoduy, and M. Keyvan-Ekbatani, Integrated deep learning and stochastic car-
11       following model for traffic dynamics on multi-lane freeways. *Transportation research part*
12       *C: emerging technologies*, Vol. 106, 2019, pp. 360–377.
13 23.   Lee, S., S. C. Wong, C. C. C. Pang, and K. Choi, Real-Time Estimation of Lane-to-Lane
14       Turning Flows at Isolated Signalized Junctions. *IEEE Transactions on Intelligent Trans-*
15       *portation Systems*, Vol. 16, No. 3, 2015, pp. 1549–1558.
16 24.   Lee, S., S. Wong, and Y. Li, Real-time estimation of lane-based queue lengths at isolated
17       signalized junctions. *Transportation Research Part C: Emerging Technologies*, Vol. 56,
18       2015, pp. 1–17.
19 25.   Ngoduy, D., S. Lee, M. Treiber, M. Keyvan-Ekbatani, and H. Vu, Langevin method for
20       a continuous stochastic car-following model and its stability conditions. *Transportation*
21       *Research Part C: Emerging Technologies*, Vol. 105, 2019, pp. 599–610.