Machine learning aided management of motorway facilities using single vehicle accident data

Kaewunruen, Sakdirat; Alawad, Hamad Ali H; Omura, Takashi; Saito, Motoyasu

Machine learning aided management of motorway facilities using single vehicle accident data

Sakdirat Kaewunruen [1,*], Hamad Alawad [1] , Takashi Omura [2] , and Motoyasu Saito [3]

[1] Department of Civil Engineering, School of Engineering, The University of Birmingham, Birmingham B15 2TT, UK; HXA724@student.bham.ac.uk (H.A.); s.kaewunruen@bham.ac.uk (S.K.)
[2] Chief engineer, Road Safety Service Division, Road Traffic Information and Safety Office, Maintenance and Traffic Management Department, Metropolitan Expressway Company Limited, 1-4-1 Kasumigaseki, Chiyoda-ku, Tokyo, Japan; PH +81 (0) 335399504; FAX +81 (0) 335025676; email: t.omura85@shutoko.jp
[3] Senior division manager, Road Safety Service Division, Road Traffic Information and Safety Office, Maintenance and Traffic Management Department, Metropolitan Expressway Company Limited, 1-4-1 Kasumigaseki, Chiyoda-ku, Tokyo, Japan; PH +81 (0) 335399500; FAX +81 (0) 335025676; email: m.saito1372@shutoko.jp
***** Correspondence: s.kaewunruen@bham.ac.uk

Abstract: Management of expressway networks has been mainly focused on defect management without looking at the correlations with accidental risks. This causes unsustainability in expressway infrastructure maintenance since such defects may not be a contributing factor towards public safety. Thus, it is necessary to incorporate accidental events for decision-making in infrastructure management. This study has developed a novel approach to machine learning (ML) that incorporates actual primary data from the last 10 years of single-vehicle accidents by collisions with motorway facilities (SVA) or so-called single-vehicle collisions with fixed objects. The ML is firstly aimed at identifying the influential factors of SVA in relation to finding the effective countermeasures for accidents by integrating the correlation analysis, multiple regression analysis and machine learning techniques. The study reveals that wet pavement conditions have a significant effect on SVA. The results show that improvement of the skid resistance is the most effective method to reduce SVA when the average vehicle speed (AVS) is less than 60 km/h. At the locations with gentle curve radii, ML indicates that it is crucial to redesign the speed-through management. Interestingly, the real data over 10 years indicates no relationship between equivalent single axle load (ESAL) and skid resistance, although many other studies have demonstrated the inverse relationship. In this study, the novel ML mean demonstrates excellent capability in providing suitable countermeasures for a reduction of SVA under a variety of uncertain and road quantitative aspects. The ML-based mitigation policies can also be applicable to other motorways and can contribute to their road safety, underpinning sustainable transport systems.

Keywords: safety management; risk management; sustainable maintenance; single vehicle; accidents; uncertainty; expressway; motorway.

---

1. Introduction

Metropolitan Expressway Company Limited “MECL” was established in 1959 with the aim of reducing traffic congestion in and around the Tokyo area. To reach their goals they have taken a number of countermeasures for reducing heavy traffic congestion such as expressway network expansion, introduction of the electric toll collection system and provision of correct traffic information. Despite efforts and a significantly decreased amount of traffic congestion, road safety on the expressway remains a significant issue. The expressway has experienced a significant number of accidents since it opened in 1962. In fact, about 1 million vehicles use the expressway daily and around 30 accidents still occur every day.

Certainly, road safety can be closely related to traffic congestion, as explained by Li et al. [1], when they describe a strong correlation between traffic congestion and the probability of a rear-end collision. Thus, it was expected that through alleviating the traffic congestion this would reduce the number of traffic accidents [2]. However, it has been found that one of the major causes of accidents is speeding and therefore, reducing traffic jams alone is not enough to reduce accidents. Despite the unclear relationship between the mean speed and the accident rates, it is imperative to decrease the number of SVA accidents and their consequences in order to positively affect the sustainable development of a society. The traffic accidents lead to economic losses such as medical expenses, loss of production and damage of vehicles and road facilities [3][4]. Statistically, in the Metropolitan Expressway in Japan, the total number of accidents has not really changed in the last ten years, as shown in Figure 1.

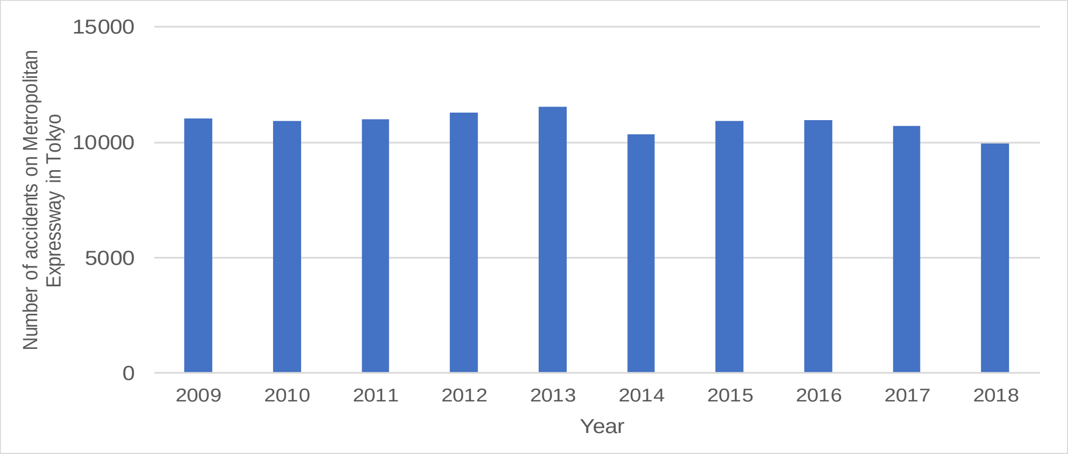


Figure 1.Transition of the number of accidents on Tokyo Metropolitan Expressway.

The literature has established that SVA and multiple-vehicle collisions have a wide variety of variables, influences, and different circumstances [5]. It is also clear that few studies have been conducted to examine SVA [6][7]. Moreover, for transportation infrastructure systems under the various hazards, the details of SVA can be a safety performance and resilience indicator of the traffic flow on the highways [8].

In addition, there is very limited research investigating single-vehicle accidents (SVA) by collisions with motorway facilities or so-called ‘single-vehicle collisions with fixed objects’ in Tokyo. MECL is seeking effective road safety methods for SVA in Japan, about which there has been very little research conducted so far. Therefore, this study focuses on SVA to fill the research gap and provide the Japanese transport industry with some important insights. Currently, data show that SVA tends to happen on sharp curves [9], with most SVA in Metropolitan Expressway happening there in recent years. Therefore, this study aims to identify factors influencing SVA with regard to quantitative influential factors as well as the influential factors of uncertainty such as weather conditions, traffic volume, vehicle speed, and skid resistance of a pavement. As for the skid resistance, the pavement condition based on skid resistance of a pavement has been proven to be closely linked to the road accident rate [10]. However, MECL has not used skid resistance as the criteria of a pavement reconstruction because the relationship between safety risk and the skid resistance has not been identified on the Metropolitan Expressway. Hence, this research aims to focus on the relationship between the skid resistance and SVA as well as to identify the relationships among other conceivable traffic accidental factors demonstrated above. Furthermore, 30 locations, which have the biggest number of SVA over the last 10 years, are focused on in the analysis. The scope of this study is to identify the influential factors behind SVA (see Figure 2).

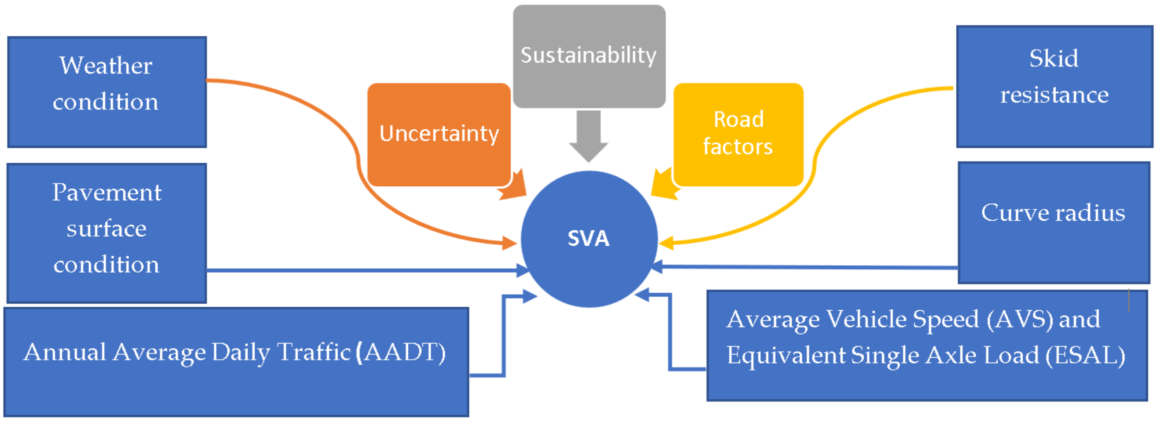


Figure 2.The influential factors causing SVA.

In addition, the study focuses on both quantitative road factors and uncertainty parameters. It aims to investigate countermeasures corresponding to each factor:

- to identify the relationship between the influential factors and SVA from the viewpoint of both uncertain and quantitative road factors, and;
- to investigate countermeasures for reducing the risk of SVA.

Since accidents involve complex interaction factors, novel techniques are required for better analytics, including predictions and supporting real-time decisions utilizing ML.

Statistical models are designed for inference about the relationships between variables, and ML is designed to make the most accurate predictions possible in order to obtain a general understanding of the data to make predictions.

The ML has been proven to deliver more accurate analysis data than the traditional methods, and it can deal with many dynamic factors in real-time when compared to statistical (regression) models. The ML models can train and can be used for predictions, engineering redesign, and advanced analytics in order to enhance safety and reduce SVA.

The outcome of this study will support decision-making processes in order to prioritize maintenance and repair activities of the motorway facilities on the Tokyo Metropolitan Expressway, Japan.

2. Literature Review

A number of approaches to reduce traffic accidents have been adopted from various perspectives. At present, it has become apparent that road safety factors can be mainly divided into three categories, such as human factors, vehicle factors, road and various environmental factors [11][12][13][14][15]. Although most problems related to vehicle factors have already been tackled, other problems associated with human and road environmental problems still exist in Japan [16][17]. Furthermore, more than 90% of road traffic accidents are due to human factors [18] and as a result, many studies have focused on human behaviours. Nishiuchi [19] [20] and Hung and Huyen [21] found that legislation, enforcement and education were effective for the reduction of road traffic accidents. In addition, Rolison et al. [22] demonstrated that road safety risk increases when law enforcement practices are inadequate, according to expert views and accidental records. They concluded the inadequacies could cause drivers' carelessness, which could cause road traffic accidents. Moreover, it has been found that the most influential seven risk factors are related to human factors [23]. Many countries have tried to tackle the issues by using measures related to those seven risk factors. In fact, the number of accidents from driving under the influence of alcohol, which is one of the seven risk factors, has fallen in Japan by means of increasing penalties. However, the number of road traffic accidents caused by drink-driving and also distracted driving still remains high despite the fact that the Japanese government has also taken countermeasures from an educational point of view [20][24]. Therefore, it is assumed that the countermeasures against road traffic accidents from other points of view need to be considered at the same time. Another unsolved road safety factor is the "road environmental factor", which is also considered in this study. Uchida et al. [25] and Buss et al. [26] stated that road accidents related to human factors can be caused by the interaction with uncertain road environmental factors for drivers such as weather conditions and traffic situations. In other words, to identify a way to reduce road traffic accidents in terms of uncertain roads, environmental factors could reduce accidents linked with human-related factors. Jung et al. [27] revealed vehicle to vehicle crashes tend to occur on rainy days under certain conditions such as places where pavement surface material changes. In addition, Malin et al. [28] demonstrated the risk in poor weather and road conditions were higher on motorways compared to general roads, which are two lane or multiple lane roads, although the overall risk was lower on motorways. In addition, Üzümcüoğlu et al. [29] showed in their study that the traffic situation was closely related to driver behaviors; and it was found that these might be critical aspects in road safety. Moreover, quantitative road environmental factors such as road geometry and alignment also could be influential to road traffic accidents, much like uncertain road environmental factors such as weather conditions and traffic situations. Dadashova et al. [30] demonstrated that geometrical design factors such as narrow lanes, higher super-elevation, steeper slope and curve radius were found to contribute to the severity of the accident. Furthermore, Papadimitriou et al. [23] focused on road traffic accidents with road facilities as the risk factor in accidents. They then categorized road-infrastructure related crash risk factors based on how detrimental they are to road safety in consideration of safety risk level. Furthermore, Yan et al. [31] analyzed characteristics of rear-end accidents by the use of correlation analysis as well as multiple regression model, which can be suitable for this kind of research. The study also revealed seven influential road environmental factors, five factors related to the striking role and four factors related to the stuck role as the significant causation of rear-end accidents. As described above, both uncertainty factors and road quantitative aspects could influence road traffic accidents. Moreover, there have been very few studies, which focus either on SVA or on road traffic accidents that have occurred on city expressways. The correlation analysis to determine the relationship between road traffic accidents and the influential factors has been reported to be effective and suitable for this research. Hence, this study aims to identify the meaningful causation of SVA from the viewpoint of uncertainty parameters and road quantitative factors by the adoption of correlation analysis.

3. Methodology

The 10-year data sets, which include the number of SVA, AVS, Annual Average Daily Traffic (AADT) and skid resistance, are provided by MECL. This section describes methodologies utilized in the analysis.

3.1 Pearson's correlation coefficient analysis

Pearson's correlation coefficient generally shows the connection between two continuous variables. It is defined as the ratio of the covariance of the two variables to the product of their respective standard deviations, commonly expressed by "ρ". Pearson's correlation coefficient is illustrated in Equation (1).

$$\rho = \frac{Cov(x,y)}{\sigma_x \sigma_y} \quad (1)$$

The sample correlation coefficient "r" can be obtained by applying the sample covariance and the sample standard deviations into Equation (2).

$$r = \frac{\sum_{i=1}^{n}((x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2 \sum_{i=1}^{n}(y_i-\bar{y})^2}} \quad (2)$$

where:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}, \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

Pearson's correlation coefficient ranges from -1 to +1. If ρ is more than 0, two variables tend to increase or decrease simultaneously, which means positive monotonic association. Furthermore, if ρ is less than 0, one variable tends to increase when the other decreases, which means negative monotonic association. If ρ is 0, it corresponds to the absence of the monotonic association, or there is no association in the case of bivariate normal data [32]. In addition to that, the value of ρ indicates the strength of the monotonic relationship between the two variables. ρ of 1 indicates a complete linear relationship.

This study uses the analysis method to identify the relationship between the number of SVA and both uncertainty and quantitative road factors. These factors are independent so this method can be fully adopted. In this case, annual data are used for the analyses of correlation coefficients.

3.2 Multiple regression analysis

In principle, multiple linear regression is a simple extension of linear regression. However, instead of relating one dependent outcome variable y to one independent variable x, one tries to explain the outcome value y as the weighted sum of influences from several multiple independent variables shown in Equation (3).

$$\mathrm{y} = \mathrm{k} + \mathrm{a}x_1 + bx_2 + cx_3 + \cdots + \varepsilon \tag{3}$$

k illustrates the intercept of the line on the y-axis. a, b, and c are the slopes of the relations between y and x1, x2, and x3, respectively. Moreover, "ε" shows the random error term. Basically, this equation plots the best fitting line. However, it is plotted through n + 1-dimensional space [33].

Multiple regression models are generally harder to yield best-fitting than single-parameter linear regression models because different independent variables may not be independent of each other (Cohen et al., 1983). Furthermore, independent variables need to be disentangled from each other mathematically to optimize the multiple regression equation [33]. Therefore, a multiple regression model is fitted by throwing out independent variables which have no significant relation to y and by normalizing independent variables in a manner that removes how they are influenced by other variables. As a result, a multiple regression model can be obtained. The final fitted model can explain the amount of an increase in one unit in each independent variable. Although this analysis method is a linear approximation and the equation may not fully or accurately account for the relationships between variables and outcomes, in practice, linear models are generally examined at first by how well they perform before considering more complicated nonlinear regression methods. It is noted that MECL has not analyzed the relationship between SVA and the aspects focused on in the study before. Thus, this study is imperative to identify the relationships among relevant factors and the combination of the factors, which are likely to impact SVA. Moreover, for the smart analytics of AI, ML analysis and decision tree methods that support nonlinearity have been considered, and some techniques are compared in Section 9 Machine learning analysis.

4. Data

This study focuses on SVA and thus, SVA observation data are essential for the analysis. The research makes use of data sets from 30 locations where the most frequent numbers of SVA occurred on the Metropolitan Expressway between 2009 to 2019. Those places are shown in Figure 3 and Table 1.

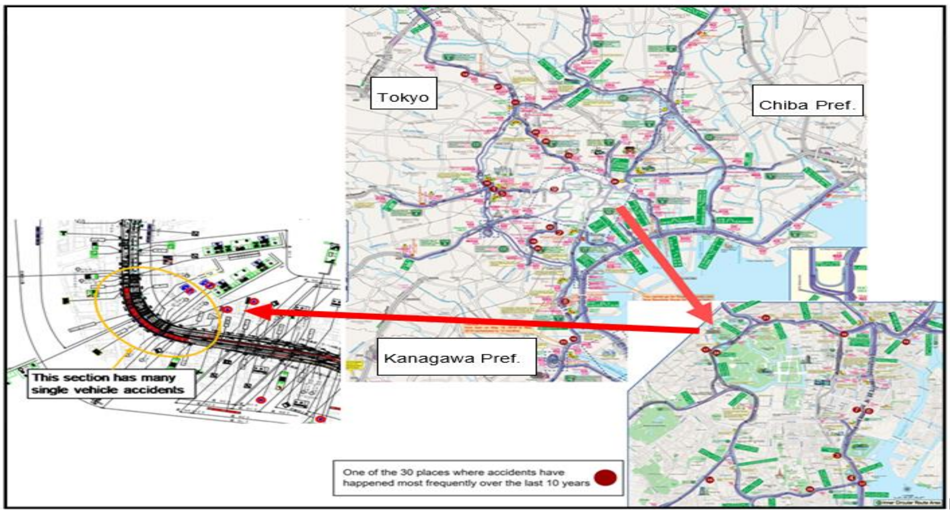


Figure 3. 30 places where accidents have happened most frequently over the last ten years on the Metropolitan Expressway with the example of a curve section.

Table 1. 30 places where accidents have happened most frequently over the last 10 years.

Number	Route	Direction	Start km post	End km post	Name of the area
1	Route 4 Shinjuku Line	Inbound	4.8	5.3	Sangubashi curve
2	Route 2 Meguro Line	Outbound	1.1	1.6	Furukawabashi curve
3	Inner Circular Route	Inner	3.5	4.0	Shiodome S curve, Shiodome JCT
4	Inner Circular Route	Inner	4.2	4.7	Hamasakibashi JCT
5	Route 4 Shinjuku Line	Outbound	4.3	4.8	Yoyogi S curve
6	Inner Circular Route	Outer	2.5	3.0	Ginza S curve
7	Inner Circular Route	Outer	2.4	2.9	Ginza curve
8	Route 1 Haneda Line	Outbound	3.6	4.1	Tennozu S curve
9	Route 4 Shinjuku Line	Inbound	0.9	1.4	Benkeibori curve
10	Route 1 Haneda Line	Inbound	11.4	11.9	Haneda tunnel
10	Central Circular Route	Outer	25.0	25.5	Itabashi JCT
12	Inner Circular Route	Outer	4.4	4.9	Hamasakibashi JCT
12	Route 5 Ikebukuro Line	Inbound	2.3	2.8	Omagari curve
12	Route 5 Ikebukuro Line	Outbound	14.1	14.6	Nakadai S curve
15	Route 2 Meguro Line	Outbound	0.4	0.9	Ichinohashi JCT
16	Route 6 Mukojima Line	Inbound	0.1	0.6	Hakozaki JCT - Edobashi JCT
17	Inner Circular Route	Outer	11.0	11.5	Chiyoda tunnel, Sanbancho curve
18	Yaesu Route	Southbound	0.6	1.1	Yaesu tunnel
19	Route 2 Meguro Line	Outbound	4.0	4.5	Osaki curve
20	Route 5 Ikebukuro Line	Inbound	4.7	5.2	Gokokuji S curve
21	Inner Circular Route	Outer	12.6	13.1	Takebashi JCT - Kandabashi exit
21	Route 2 Meguro Line	Inbound	0.8	1.3	Furukawabashi curve
23	Route 1 Haneda Line	Inbound	12.4	12.9	Haneda curve
24	Inner Circular Route	Inner	10.9	11.4	Chiyoda tunnel, Sanbancho curve
25	Route 2 Meguro Line	Outbound	5.0	5.5	Gotanda S curve
26	Route 1 Ueno Line	Inbound	0.0	0.5	Honcho entrance - Edobashi JCT
26	Route 5 Ikebukuro Line	Outbound	11.0	11.5	Itabashi-honcho curve
28	Route 5 Ikebukuro Line	Inbound	5.9	6.4	Hinode-daiichi curve
29	Inner Circular Route	Inner	5.4	5.9	Shibakouen S curve
30	Route 4 Shinjuku Line	Inbound	5.7	6.2	Shinjuku curve

It is notable that, uncertainties, weather conditions, traffic volume and AVS are the main focus of the analysis. In fact, Theofilatos and Yannis [34] stated that weather conditions are closely related to road traffic accidents. On the other hand, they do not analyze what kind of accidents were linked to weather conditions. Therefore, it is important for this study to find the association between weather conditions and SVA. The research focuses on pavement surface condition in terms of humidity, such as dry and wet conditions, similar to that which was studied by Theofilatos and Yannis [34]. As for traffic volume, this is known as one of the reasons for accidents involving collisions between vehicles [35]. However, previous studies have not shown the clear relationship between traffic volume and SVA, especially on the city expressway. Thus, this study analyses the relationship between SVA and traffic volume by use of Annual Average Daily Traffic (AADT). Moreover, AVS is utilized for the analysis. Generally speaking, the vehicle speed is widely believed to be a key issue in the cause of road traffic accidents. Tanishita and Wee [4] focused on reducing speed on the rural expressway in Japan and they demonstrated the highest probability of an accident occurring is when speed reduces from 110 to 85 km/h. However, speed limitations on the rural expressway and the city expressway are quite different. Therefore, this study aims to identify the relationship between the vehicle speed on Metropolitan Expressway (which is the city expressway) and SVA. This study uses AVS as a criterion for expressing vehicle speed each year in each location. For the skid resistance of a pavement, although the relationship between the skid resistance and the risk of road accidents are tangible, as mentioned before, a measurement has not been conducted on the Metropolitan Expressway for many years. The Metropolitan Expressway was originally built to secure the transport capacity for the Tokyo Olympic Games in 1964. At that time, most parts of it were constructed over small water channels and limited public land. As a result, there are many sharp curves as well as many branching and merging points in a short section. Thus, MECL has placed heavy emphasis on countermeasures for road safety in consideration of the road alignment. They did not, however, consider the skid resistance. More recently, MECL has started to focus on the skid resistance to reduce SVA because those accidents have become more serious in recent times. Although the data for skid resistance have not been measured in whole road sections of the expressway, the data have been measured in places where SVA often occurs. The measurement frequency ranges from once a year to once every 10 years. In this study, 79 sets of data for the skid resistance of a pavement are obtained. To determine the effect of quantitative road factors, road curve radius is utilized. In addition, all locations in focus are curved sections, as demonstrated in Figure 4. In fact, most SVA happens in curve sections. Thus, in this part of the analysis, each curve radius in each curved section is determined. Table 2 provides general information for the input data used in the analysis, the raw data has been cleaned for analysis, and there are 30 locations and 10 years of data giving 300 data points. In addition, AADT, AVS, skid resistance and curve radius are set as independent variables; and the number of SVA is set as an explanatory variable.

Table 2. Summary of the data focused on in the study.

Input data	Number of data	Mean	Minimum	Maximum	Standard deviation
Number of SVA (accidents/year)	300	13	0	309	58
AADT (vehicles/day)	300	35,600	2,200	69,700	14,800
AVS (km/h)	300	58.2	25.7	80.6	12.3
Skid resistance (coefficient of friction: non-unit)	79	0.38	0.02	0.63	0.10
Curve radius (m)	300	118	66	309	58

5. Data Analysis and Results

This section contains the analyses of the input data by using Pearson's Correlation analysis method. This study focuses on four (4) different types of independent variables such as AADT, AVS, skid resistance and curve radius. Table 3. demonstrates the correlation coefficient of each other variable. Generally, if the absolute correlation coefficient between two different variables is more than 0.8, then multicollinearity is a problem [36]. Thus, there is low possibility of multicollinearity occurring in the case of using those four (4) types of variables.

Table 3. Correlation coefficient of four (4) types of independent variables focused on in the study.

	AADT	AVS	Skid resistance	Curve radius
AADT	-	-0.14	0.15	0.08
AVS	-	-	-0.18	0.33
Skid resistance	-	-	-	0.27
Curve radius	-	-	-	-

The correlations of the data in terms of uncertainty parameters, which include weather conditions, AADT, AVS and skid resistance, with the number of SVA, are analysed. This can be seen in Figure 4, which shows the overlapping of factors and the complexity level between them.

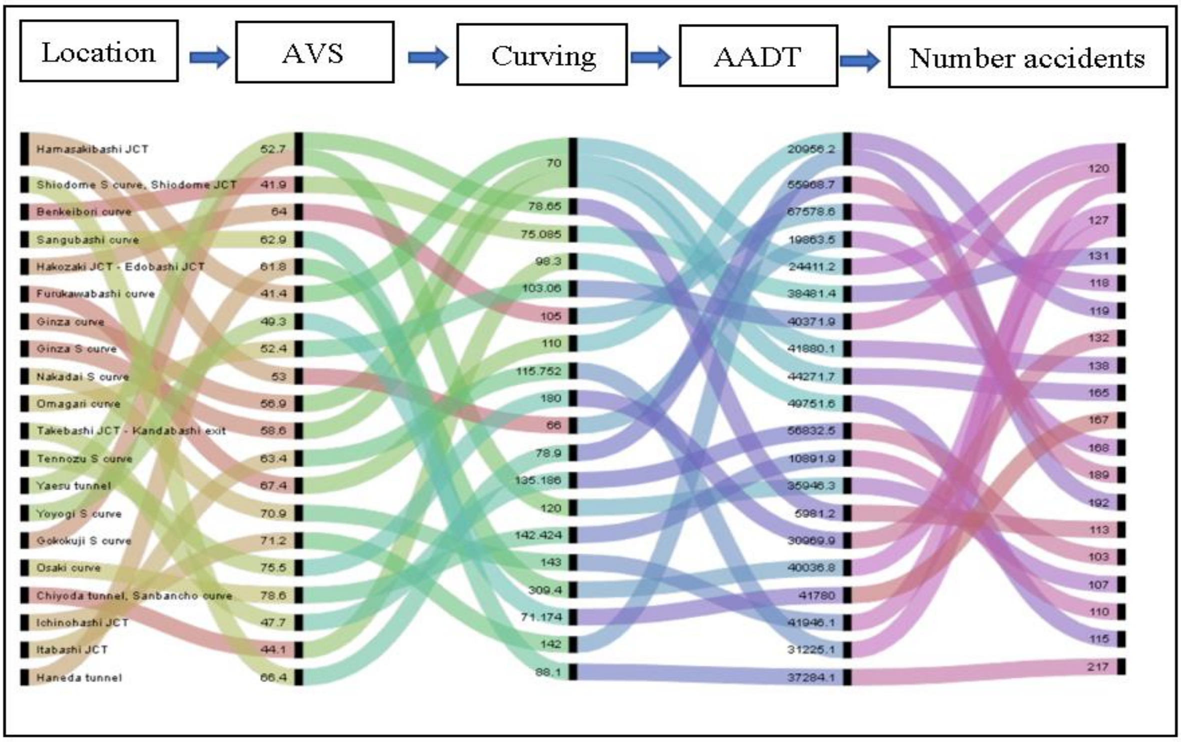


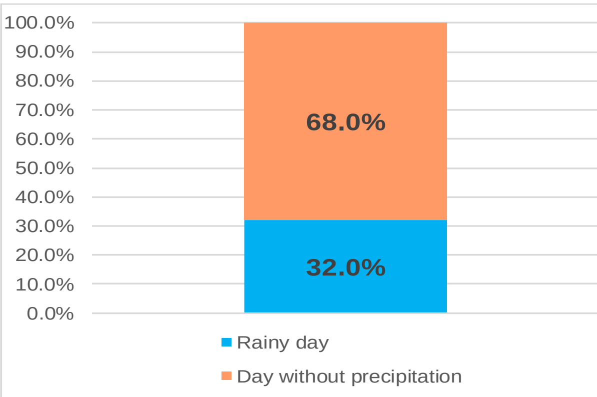
Figure 4. This visualization alluvial diagram shows examples of the data set and the complexity between the factors that need to be analyzed.

First of all, for weather conditions, Figure 5 shows the ratio of rainy days and days without precipitation from 2009 to 2019, according to Japan Meteorological Agency [37]. In addition, Figure 6 demonstrates the ratio of three (3) types of road surface conditions at the time of accidents for the analyzed period. According to those data, rainy days account only for about 30% of a year, however, the road pavement surface in wet conditions dominates SVA by more than 75%. This aligns with a study by Theofilatos and Yannis [34], which stated that road traffic accidents are more likely to happen on rainy days than on the days without precipitation. Therefore, it is noted that the accidents are closely related to weather conditions. This study then focuses on wet pavement conditions related to rainy days and also dry pavement conditions, which means days without precipitation, in order to benchmark the relationships between SVA and both uncertainty parameters and road condition aspects. With respect to the traffic volume, Figure 7 shows the relationship between AADT in each place and the number of SVA. Using the linear approximation best-fit in Figure 8, the number of SVA appears to climb with an increase in AADT. However, the correlation is quite weak. The correlation coefficient between those two aspects is almost 0. However, in the case of dry pavement surface conditions, the correlation coefficient increases to 0.3, as shown in Table 4 and Figure 8.

Regarding the influence of ESAL, Figure 9 shows the relationship between ESAL in each place and the number of SVA (Number of single-vehicle accidents per year/place), which displays the number of SVA as a meagre increase with the increase in ESAL, but this link is not strong evidence, at least at this stage of data. In addition, the factors of pavement surface conditions (wet /dry) data do not produce an adequately clear relationship (see Figure 10).

In terms of the vehicle speed, although there is an unclear relationship with SVA demonstrated in Figure 11, a positive correlation between AVS and the number of SVA under wet pavement conditions can be observed, as demonstrated in Figure 12. Considering the skid resistance of a pavement surface, there appears to be a negative correlation between the skid resistance and the number of SVA, as illustrated in Figures 13 and 14, and also the increase between ESAL and skid resistance in Figure 15. In other words, the number of SVA reduces with an increase in the skid resistance, especially in wet pavement conditions.

In addition, the curve radius can be used to represent the road quantitative factor, which is a physical parameter of road conditions. Figure 16 shows the relationship between the curve radius and the number of SVA in each location. Although there is very little connection between those two factors in Figure16, a negative correlation can be observed to some extent in the case of dry road surfaces shown in Figure 17. In order to understand the relationships mentioned above, Table 4 demonstrates the correlation coefficients between the number of SVA and each conceivable factor. Although some correlations are clear in some cases, it is difficult to determine a clear association, since the biggest correlation coefficient is 0.3 at most. Therefore, further essential analysis is conducted by using cross-tabulation analysis, as provided in the next section.



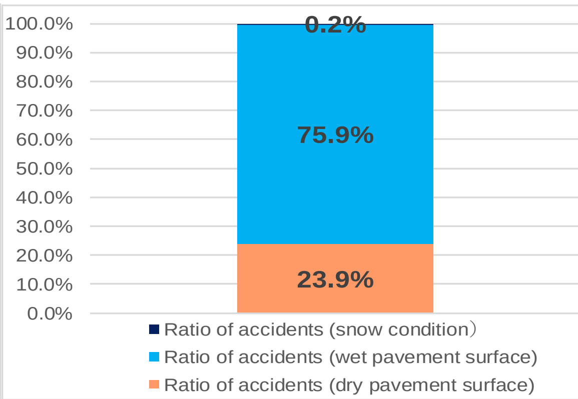
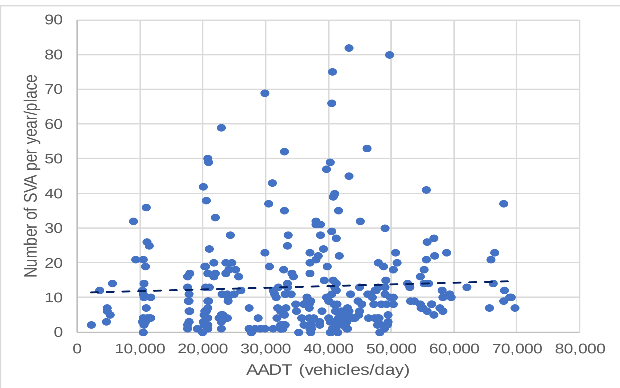
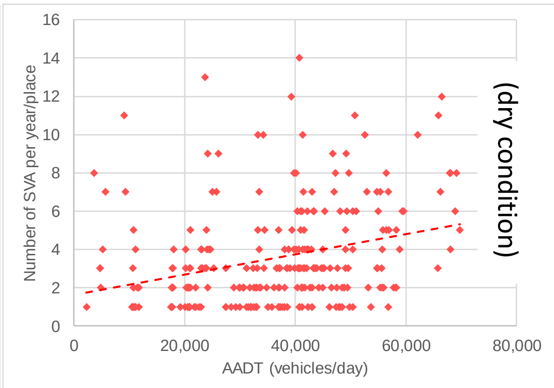


Figure 5. Ratio of rainy days and other days.

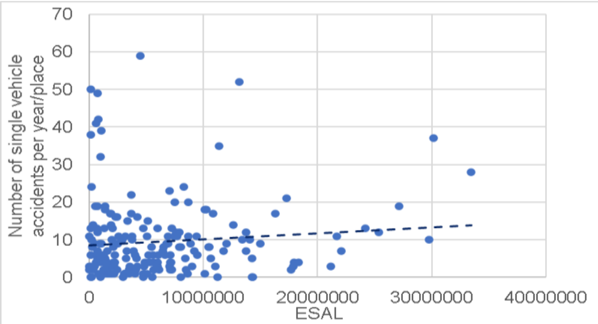
Figure 6. Ratio of each road surface condition at the time of SVA.





--

Figure 7. Relationship between AADT and the number of SVA. Figure 8. Relationship between AADT and the number of SVA



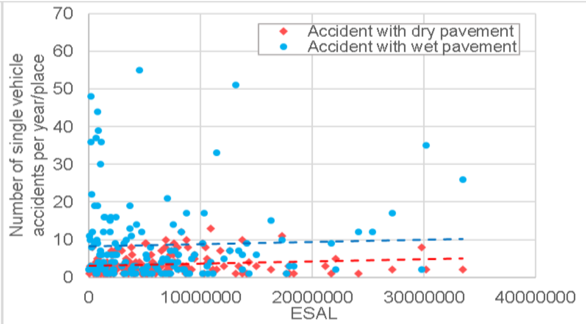
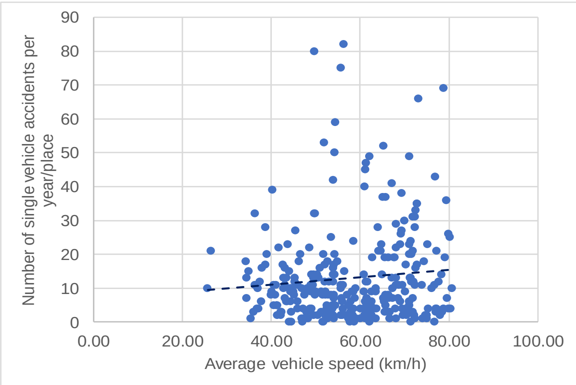


Figure 9. Relationship between ESAL and the number of SVA. Figure 10. Relationship between ESAL and the number of SVA (dry/wet) condition.



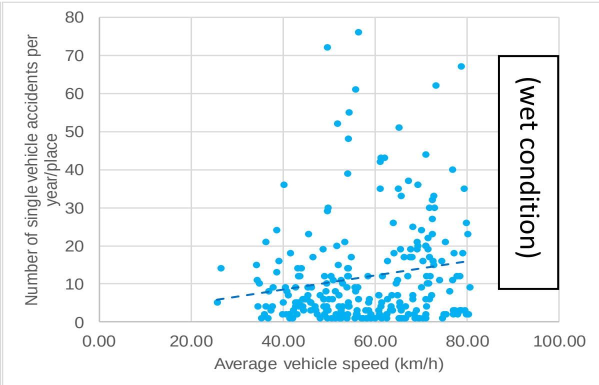
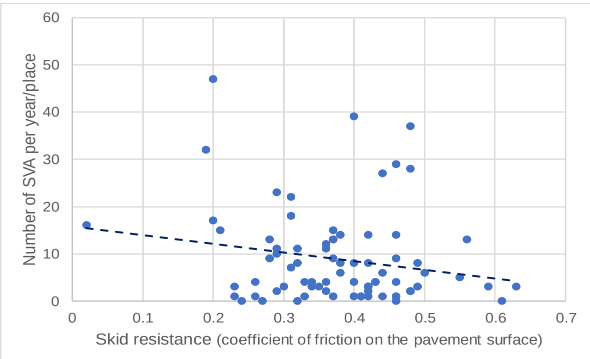


Figure 11. Relationship between AVS and the number of SVA.

Figure 12. Relationship between AVS and the number of SVA.



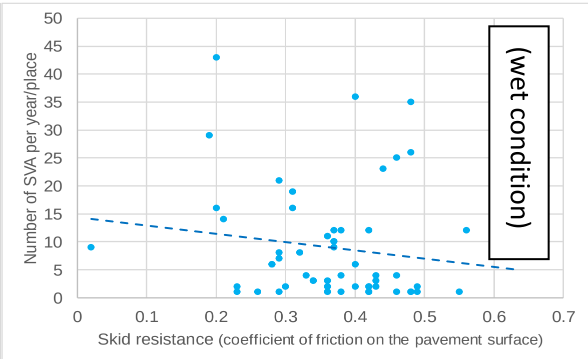


Figure 14. Relationship between skid resistance on the pavement surface and the number of SVA.

Figure 13. Relationship between skid resistance on the pavement surface and the number of SVA.

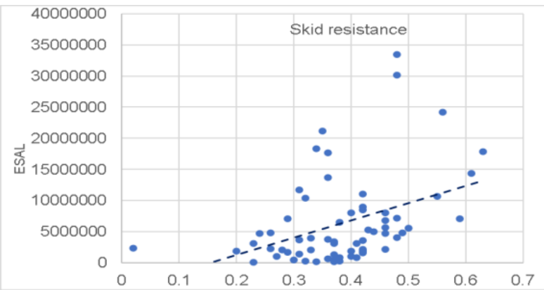


Figure 16. Relationship between Skid resistance and ESAL.

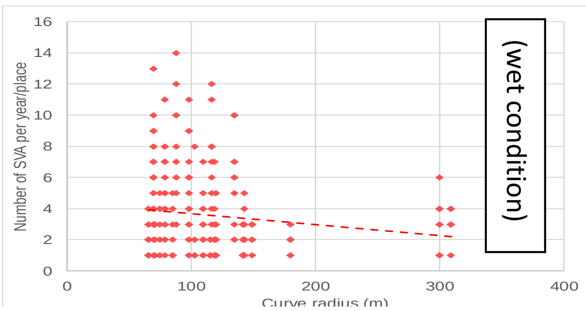


Figure 15. Relationship between curve radius and the number of SVA.

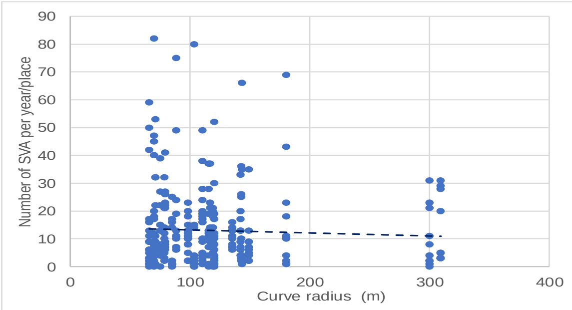


Figure 17. Relationship between curve radius and the number of SVA.

Table 4: Coefficient of correlation between the number of SVA and conceivable aspects.

SR= Skid resistance CR = Curve radius	Uncertain Aspects			Quantitative road aspect
	AADT	AVS	SR	CR
Number of SVA	0.1	0.1	-0.2	0.0
Number of SVA (Dry condition)	0.3	-0.2	0.0	-0.2
Number of SVA (Wet condition)	0.0	0.2	-0.2	0.0

6. Cross Tabulation Analysis and Results

The relationships between the number of SVA and several key factors are shown in Figures 18-25. Here, each factor can be divided into groups based on probabilistic criteria. The results without skid resistance are divided into two parts by adopting the criteria below. For the results with skid resistance, MECL has not considered such factors and there are no specific criteria for justification. Thus, the skid resistance is excluded from the group.

- The criterion of AADT
  30,000 vehicles/day can be the criterion, in 30 places focused is about 30,000 vehicles/day.
- The criterion of AVS
  60km/h can be the criterion because it is the speed limit on most routes on the Metropolitan Expressway [3][4]
- The criterion of curve radius
  Curve radius of 100m can be the criterion because MECL defines the curve section of which curve radius is less than 100m as a sharp curve, while the curved section needs to have fences, which stops every vehicle falling off the expressway.

Then, each correlation coefficient is calculated between the number of SVA and each factor categorized by the criteria defined above. As a result of the analyses, Table 5 is obtained. These analyses focus on the correlation coefficient that is more than 0.3. Mukaka [38] stated if a correlation coefficient is 0.3 after excluding outliers, it may be interpreted as a weak positive correlation. In this case, outliers are considered to be excluded after the cross-tabulation calculation is applied. Strength of correlations is shown by different colors in Table 5. In most cases, the coefficient tendencies are dependent on pavement conditions. However, only in the case of curve radius being less than 100m is the correlation between the number of accidents on both dry and wet pavements and skid resistance negative. In other words, the number of SVA reduces with the increase in skid resistance in the curved sections with a radius less than 100m. This is a new finding for MECL. The results obtained from the calculation are summarized below.

- Summarized results in **dry** pavement conditions

  <u>In the case of AVS being more than 60 km/h</u>

  ✓ The number of SVA could increase with an increase in AADT

  <u>In the case of curve radius being more than 100 m</u>

  ✓ The number of SVA could increase with an increase in AADT

  <u>In the case of curve radius being less than 100 m</u>

  ✓ The number of SVA could fall with an increase in skid resistance

- Summarized result in **wet** pavement conditions

  <u>In the case of AVS being less than 60 km/h</u>

  ✓ The number of SVA could fall with an increase in skid resistance

  <u>In the case of curve radius being more than 100 m</u>

  ✓ The number of SVA could increase with an increase in AVS

  <u>In the case of curve radius being less than 100 m</u>

  The number of SVA could fall with an increase in skid resistance

Table 5 Calc results of Pearson's correlation coefficient between SVA and each aspect.

SR= Skid resistance CR = Curve radius D= Dry surface W= Wet surface		Pavement condition	Uncertain Aspects			Quantitative road aspect
			AADT	AVS	SR	CR
Uncertain aspects	AADT (>30000 vehicles/day)	D	-	0.0	-0.1	-0.2
		W	-	0.2	-0.2	0.0
	AADT (<30000 vehicles/day)	D	-	-0.5	-0.1	-0.4
		W	-	0.2	-0.2	0.2
	AVS (+60km/h)	D	0.5	-	0.1	-0.2
		W	0.0	-	-0.1	0.0
	AVS (-60km/h)	D	0.1	-	-0.1	-0.1
		W	0.0	-	-0.3	-0.2
Quantitative road aspect	Curve radius (+100m)	D	0.6	-0.2	0.1	-
		W	-0.1	0.3	0.1	-
	Curve radius (-100m)	D	0.1	0.0	-0.3	-
		W	0.1	0.1	-0.5	-

<u>*7.1 Results of correlation in dry pavement condition:*</u>

Case 1: AADT is less than 30,000 vehicles per day.

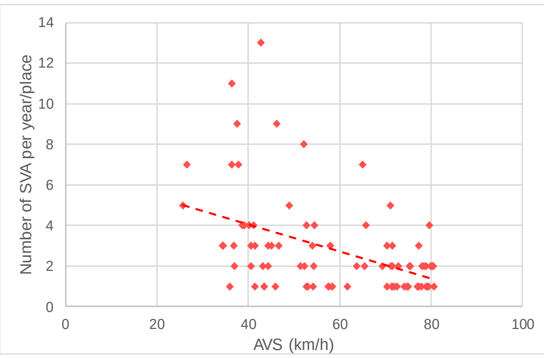


Figure 18. Relationship between AVS and the number of SVA with AADT < 30,000

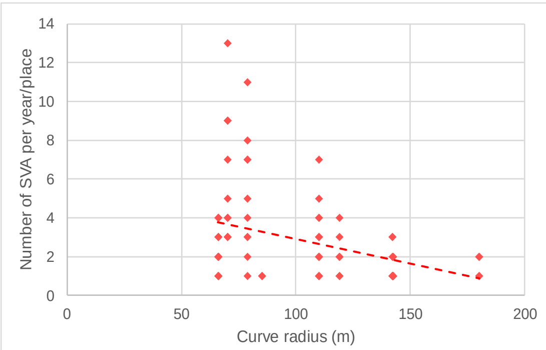


Figure 19. Relationship between curve radius and the number of SVA with AADT < 30,000

Case 2: AVS is more than 60 km/h

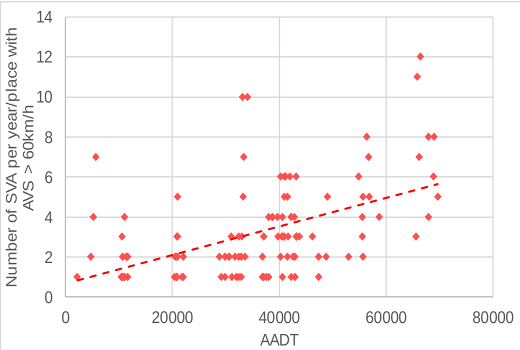


Figure 20. Relationship between AADT and the number of SVA with AVS +60km/h

Case 3: Curve radius is more than 100 m

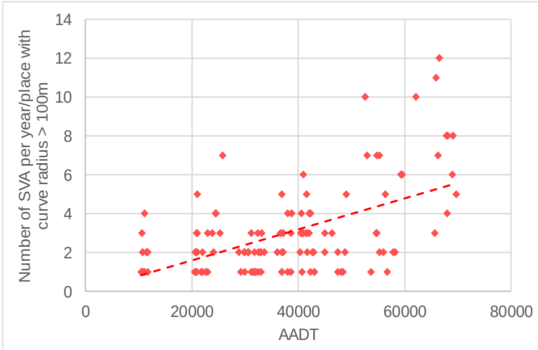


Figure 21. Relationship between AADT and the number of SVA with curve radius > 100m

Case 4: Curve radius is less than 100 m.

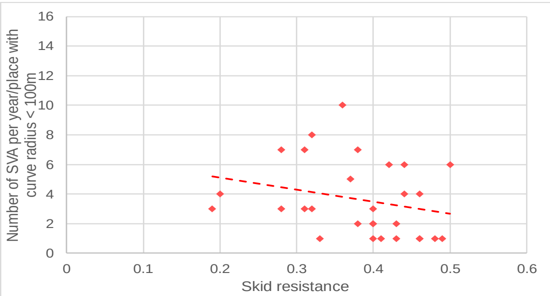


Figure 22. Relationship between the skid resistance and the number of SVA with curve radius < 100m

7.2 Results of correlation in wet pavement condition

Case 5: AVS is less than 60 km/h

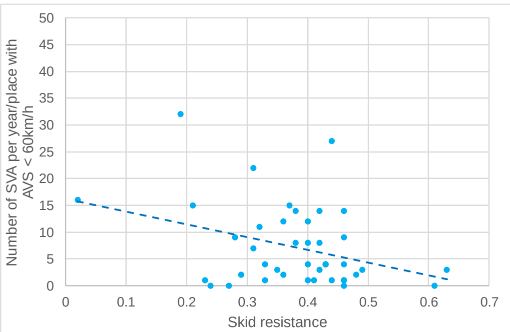


Figure 19. Relationship between the skid resistance and the number of SVA with AVS < 60km/h

Case 6: Curve radius is more than 100 m

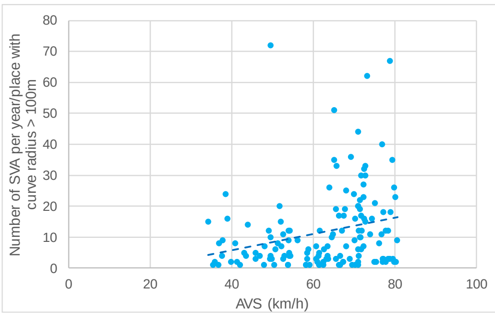


Figure 18. Relationship between AVS and the number of SVA with curve radius > 100m

Case 7: Curve radius is less than 100 m

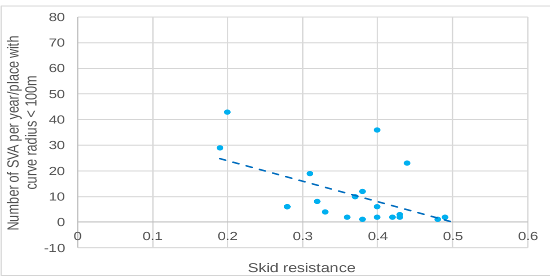


Figure 20. Relationship between the skid resistance and the number of SVA with curve radius < 100m

7. Multiple regression analysis and results

The effect on SVA of the combination of factors has been identified using multiple regression analyses. The previous section has determined Pearson's correlation coefficients between SVA and each factor. It is found that the relationship between SVA and each factor has become increasingly clear. This section focuses on the factors, which have been confirmed to have a correlation with SVA through Pearson's correlation coefficients. Firstly, the variable combination of the correlation coefficients which are more than 0.5, as shown in Table 5, is analyzed. Then, Table 6 can be obtained. The figure illustrated in Table 6 shows multiple regression coefficient R-squared "R^2" and the independent variables used in the calculation. Nau [39] stated that, "if the dependent variable is a properly stationarized series, then an R-squared of 25% may be quite good". Of course, the required size of an R-squared depends on the variable with respect to a measurement. According to Table 6, most of the R-squared values are more than 25%. However, it is not obvious whether there is a definite correlation between SVA and the variables. Therefore, this section aims to carry out a further analysis by using 3 different independent variables with the consideration of previous results, as shown in Table 5. Moreover, the correlation coefficients, which are more than 0.3, are focused. As a result of the calculation, Table 7 can be obtained. Considering the dry pavement condition, each R-squared value is less than the previous result shown in Table 6. On the other hand, R-squared values in wet pavement conditions are relatively high. The results are clearly shown in Table 8 and 9. If a p value less than 0.05 has been considered statistically significant, AVS is the most influential factor according to a t-statistics and a p value in Table 8, then, the skid resistance is the most influential factor in Table 9. This result is new to MECL and will influence the maintenance policy in Tokyo. Further interpretation of the analysis results can be seen in Table 10.

Table 6: Calculation result of multiple regression analysis with 2 independent variables.

Pavement condition	*Variable condition*				*Variable*			
	AADT	AVS	SR	CR	AADT	AVS	SR	CR
Dry	<30,000 vehicles/day	_	_	_	_	**R^2=0.23**	_	_
	_	+60 km/h	_	_	**R^2=0.26**	_	_	_
	_	_	_	+100 m	**R^2=0.31**	_	_	_
Wet	_	_	_	-100 m	_	_	**R^2=0.28**	_

Table 7.: Calculation result of multiple regression analysis with 3 independent variables

Pavement condition	*Variable condition*				*Variable*			
	AADT	AVS	SR	CR	AADT	AVS	SR	CR
Dry	<30,000 vehicles/day	_	_	_	_	R^2=0.23	_	_
	<30,000 vehicles/day	+60 km/h	_	_	_	_	_	R^2=0.21
		+60 km/h		+100 m	R^2=0.13	_	_	_
Wet	_	-60 km/h	_	+100 m	_	_	**R^2=0.66**	_
	_	-60 km/h		-100 m	_	_	**R^2=0.41**	_

Table 8. Result of multiple regression analysis in the case of AVS -60km/h & curve radius +100m

	Coefficient	Standard error	t statistic	P value	Lower 95%	Upper 95%
Intercept	50.65	14.29	3.55	0.01	15.69	85.60
Curve radius	-0.06	0.02	-2.42	0.05	-0.11	0.00
Skid resistance	-15.80	8.47	-1.87	0.11	-36.52	4.92
AVS	0.64	0.23	2.80	0.03	1.21	0.08

Table 9. Result of multiple regression analysis in the case of AVS -60km/h & curve radius -100m

	Coefficient	Standard error	t statistic	P value	Lower 95%	Upper 95%
Intercept	33.79	35.10	0.96	0.35	-41.50	109.08
Curve rad	0.24	0.37	0.66	0.52	-0.55	1.04
Skid R	-63.80	23.67	-2.70	0.02	-114.57	-13.03
AVS	-0.40	0.34	-1.19	0.25	-1.13	0.32

Table 10. Interpretation of the analysis result

Condition		Results
AVS	Curve radius	
Less than 60 km/h	More than 100m	SVA increases with an increase in AVS
	Less than 100m	SVA reduces with an increase in skid resistance of a pavement

8. Machine learning analysis

The conventional traffic road safety accident records and analyses have opinions of experts and human interventions and this creates a lack of structure and ambiguity in the research-practice relationship [40][41]. Supporting the automation process with a decision-maker and covering more data in the system in real-time is the future direction with huge dynamic changes, including drivers' behaviors, severe weather, and smart transportation. Therefore, as a part of this study, ML has been conducted. The ML methods provide innovation trends, create models to find the accident-prone zone and different accident-factors, and mines to determine the association between these factors. The data mining is expected to overcome shortcomings produced by the Statistical analyses [42]. The data-driven methods have the potential to handle extremely large amounts of data and provide a high level of prediction accuracy y in the highway-safety analysis [43]. The techniques to be used in this study are clustering and classification, which have been used for analysis of the road accidents [44][45]. These analyses have shown the ability to help the transport authorities in improving safety requirements and recognizing the accidents' root causes [46]. Moreover, ML models have been proven as an effective analytical tool for the multiple factors associated with Road Safety [47][48].

In the beginning, cleaning and normalizing the data are carried out to run the models for predictions, where the input is set as predictors targeting one output, such as the number of accidents or predicting the locations of an accident, and the method provides many prediction paths for different factors. However, the limit of the data appears as a challenge and proves the importance of gathering more data for future technologies. This is important for learning and validation of well-balanced analytics and avoiding both overfitting and underfitting, where reliable predictions are essential. In training/ testing data split by choosing subsets for generating the 80%/20% and 60%/40% split of the dataset (two splitting data ratio), the 20% & 40% test data are not part of the training subset and are applied to predict and test the performance. For exploring the hidden patterns and analysis, the decision tree (DT) model prediction has been conducted as a machine learning (ML) theory. The decision tree has the power to simplify the data depending on the available attributes by rules to find a prediction path. Moreover, there are more selections to model and predict utilizing other predictors and many options for inputs for future research (see Figure 26).

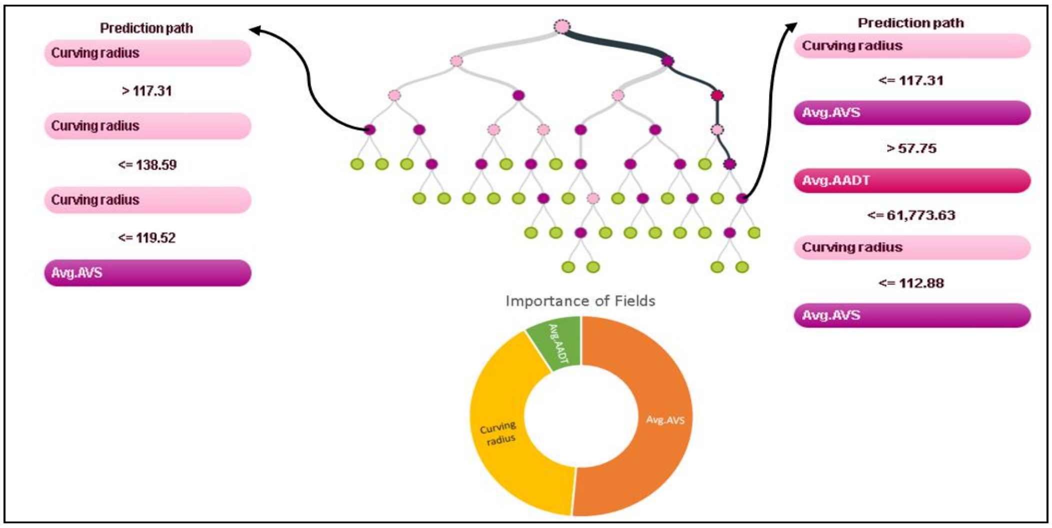


Figure 21. The DT, with an example showing the prediction paths and analyzing the data set.

The DT is secondary here for proving the power of ML in analyzing safety data which enables one to review and visualize descriptive statistics of the dataset. However, this method requires big data to obtain maximum benefit in analysis. This use of ML emphasizes one of the primary safety concepts, which is that lessons must be learned from the accidents but there must also be benefits from technology in the field. The model reaches 83.3 % and 75% accuracies for 80%/20% and 60%/40% splits, respectively, with the inbound direction as the positive class, while the data set has four rules linked to the accidents: inbound, inner, outbound, and outer (see Figure 27).

Three groups have been found to be related to AADT (<=23k,23-50k,>=50k) and to the AVS for specific locations selected as inputs for the predictions. Multiple options can be generated, and precise thresholds have been found in the complex relations which manifest the power of ML.

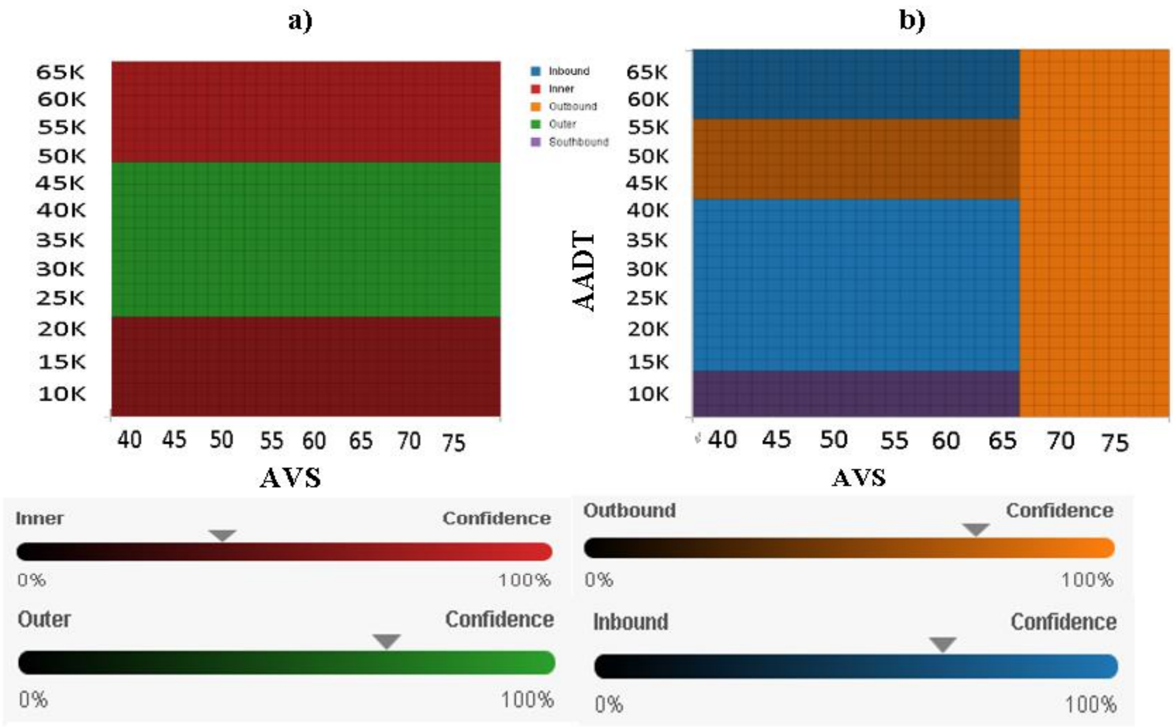


Figure 22. The Partial Dependence Plots (PDP) targeting the direction shows the marginal effect of inputs of cases on the predicted results of the ML, a) Inner Circular Route, b) Route 2 Meguro Line

There are some specific spots linking the inputs with the number of accidents. For instance, the AVS above 60 km/h and curving radius below 119 m forms an apparent effect for the number of accidents, for example, number of accidents is predicted to be 165 with curving radius 111.6m, which was a 95 with 119.2m. Also, this linked to AVS, where the speeds over 46.6 km/h have been noted as critical points for increased accidents and others highlighted points over speeds 61.4 km/h, with fixed input of AADT 30.8K.

The factor AVS and curving radius have more importance in that field and the analysis proves they are strong predictors.

Moreover, many slides of relations between the inputs can be produced for each assuming selected contributions; this method of ML is essential for designers and analyzing the safety data accurately. This analysis closely matches and confirms the previous study in that AVS has influences on SVA, and AADT and curve radius affects SVA. The sharp curves and increased AADT are factors presenting as dominating the number of accidents (see Figure 28).

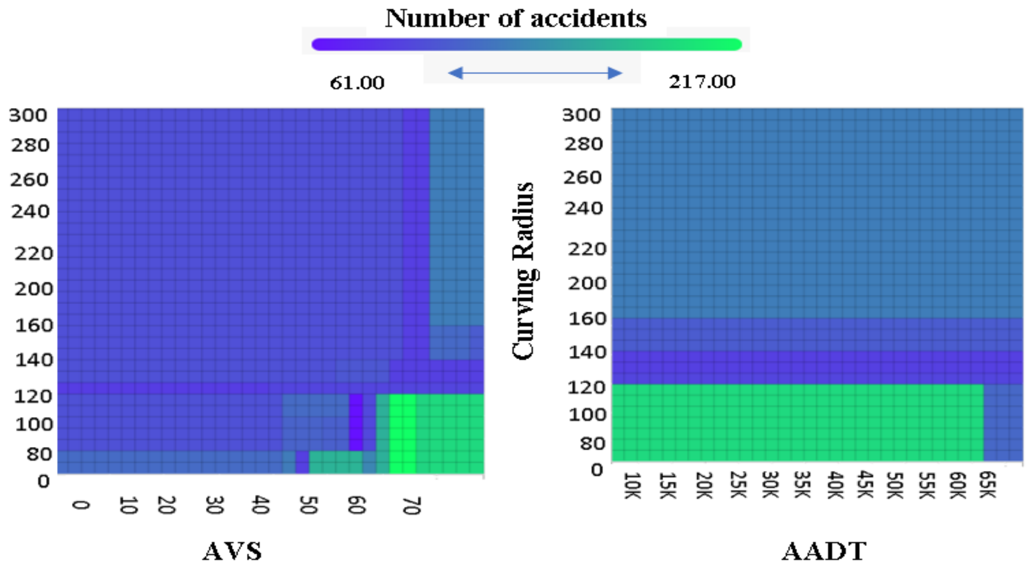


Figure 23. The Partial Dependence Plots (PDP), shows the marginal effect of inputs of cases have on the predicted results of ML, a) Input AADT = 3082.38, b) Input AVS = 58 Km/h.

For a further exposition of ML techniques in such cases, a clustering method is applied as intelligent methods used to present and extract data groups of interest are searched. It is shown that ML is a powerful analysis method for safety. To analyze the unsupervised dataset, ML is chosen with the K-means algorithm (canonical clustering), where the number of clusters is three. However, the remainder of the work is supervised ML. Utilizing cluster analysis involves separating datasets into subsets of instances (clusters), and the algorithm affirms similarities and that the groups are closer to one another if they are more similar and farther away if they are dissimilar (see Figure 29 and Table 11).

Figure 29. The 3-cluster diagram nodes and histograms.

Table 11. The details of the clusters.

Centroid name	Instances	Minimum inter-centroid	Mean inter-centroid distance	Maximum inter-centroid	Distance sum squares	Distance standard deviation	Distance sum	Distance median	Distance maximum	Distance minimum	Distance variance	Distance mean
Cluster 0	10	0.78	0.87	0.96	4.20	0.17	6.27	0.61	0.92	0.42	0.03	0.63
Cluster 1	7	0.93	0.95	0.96	3.22	0.17	4.61	0.67	0.89	0.39	0.03	0.66
Cluster 2	13	0.78	0.86	0.93	5.78	0.19	8.32	0.64	1.13	0.38	0.04	0.64

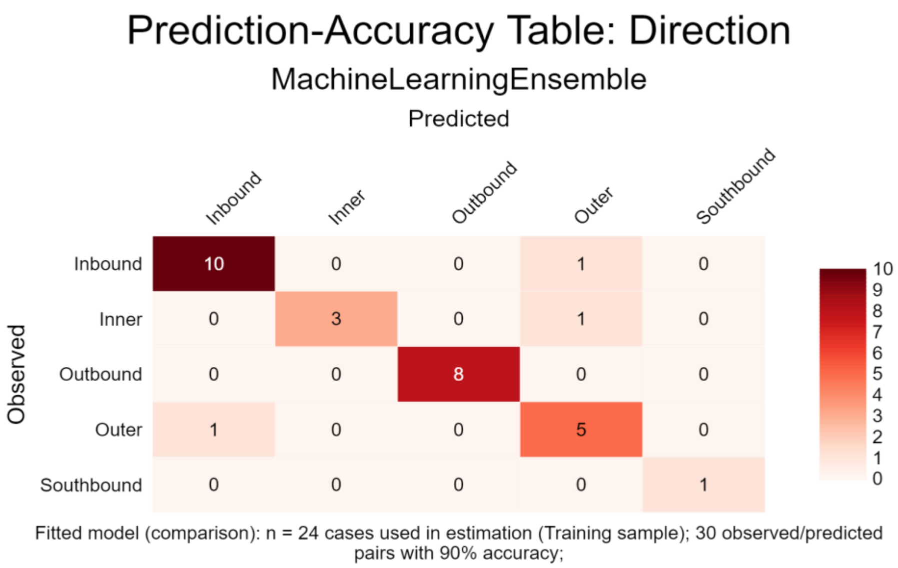
The prediction targets provide various analysis and different levels of accuracy, which present that the ML is a powerful tool for predictions and advance analytics. Thus, comparison of the models has been conducted (See Table 12). The Random Forest (RF), Gradient Boost (GB), Support Vector Machine (SVM) and optimal Ensemble respectively show better accuracy.

Table 11: The comparison of four (4) models and Ensemble with training data set and the optimal ensembles: model 2(Gradient Boost).

	Underlying model	Model type	Training RMSE	Training R^2
Model 1	RandomForest	trees = 500	0.11	0.7891
Model 2	GradientBoost	booster = gbtree	0.04	0.9791
Model 3	SupportVectorMachine	cost = 1	0.17	0.5343
Ensemble	Ensemble	All models	0.10	0.8269
Optimal Ensemble	Ensemble	Optimal	0.04	0.9791

From the comparison models, the model spots the areas correlated to the safety factors. Some areas are predicted as more affected by factors and number of accidents than others. For a diagnostic, the heat map shows the observed and the predicted values (confusion matrix) and presents the prediction accuracy (see Table 14).

Table 12: The Prediction-Accuracy table showing the observed and predicted values, as a heatmap for the ensemble model as the routes are output.



From the DT model the model is validated and the outcomes of the prediction are labelled as either positive or negative. If the prediction is positive and the actual value is also positive, then it is called a true positive (TP); with the same concepts, false positives (FP), true negatives (TN), and false negatives (FN) are realised. The four outcomes can be formed as a confusion matrix, with acceptable confidence values shown below (see Table 15).

Moreover, the area under the curve (AUC) was measured under the ROC curve. The decision tree achieves higher AUC values of 0.89 (see Figure 30).

Table 13. The evaluation results of the performance per class in the confusion matrix.

ACTUAL VS. PREDICTED	Inbound	NEGATIVE CLASS	ACTUAL	RECALL	F	Phi
INBOUND	1	0	1	100%	0.67	0.63
NEGATIVE CLASS	1	4	5	80%	0.89	0.63
PREDECTED	2	4	6	90% AVG.RECALL	Avg. F = 78 %	Avg.Phi = 0.63
PRECISION	50%	100%	75 % AVG. PRECISION	83.33% ACCUARCY	TP FN TP TN	

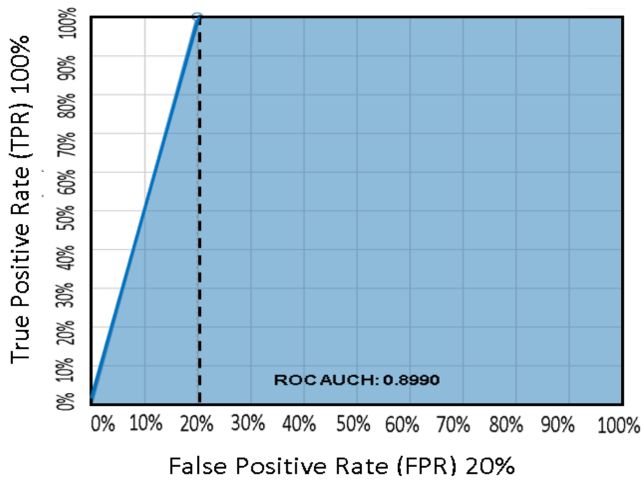


Figure 24.The ROC curve, which shows that the area under the curve (AUC) is 0.89, evaluates the model with 80% training data vs. 20% test data.

ML has beneficial and precise information presenting the ability to advance analysis compared with the traditional statistical analysis. From the LDA model, the AVS and AADT inputs have more effect on the number of accidents displayed as near points; also, the curving radius (CR) and AVS are inputs influencing some locations together (see Figure 31).

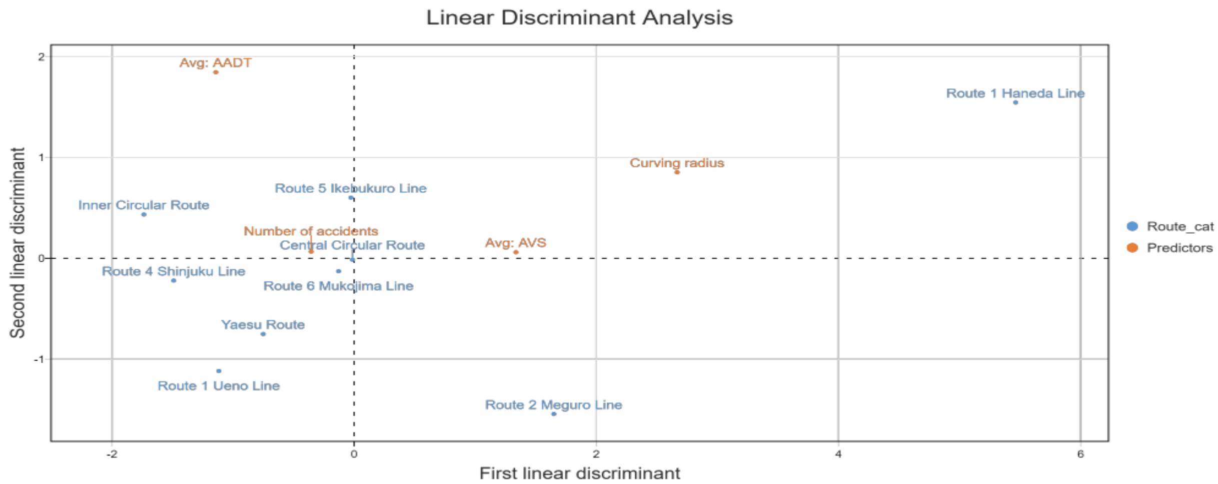


Figure 31. The Scatterplot from the LDA model targeting routes and four predictors (CR, AADT, AVS and the number of accidents).

ML is one of the industry’s future revolutions. Likewise, there is the benefit of learning and reduction of human error and intervention through configuration with the smart life (autonomic-smart cities), long term advantages that can be gained with such an application [49][50].

Finally, The ML is not just a prediction tool as it has other merits, including the visualization of data, learning and real-time analysis for assisting the decision-maker in a proactive way.

9. Discussion

This study determines influential factors of SVA and the combinations of the factors that have a significant effect on SVA. It also demonstrates that the skid resistance is the main influential factor under various conditions. Furthermore, this study determines whether there is a relationship between ESAL and skid resistance.

In addition, according to TRL [51], their pavement design manual "Oversea Road Note 31" specified Equivalent Single Axle Load (ESAL) as the pavement design factor. In fact, MECL has never tried to quantify the relationship, so this study provides new insights for MECL in order to update its practice; plus, skid resistance in the Metropolitan Expressway is not measured regularly, and ESAL is also not available on a yearly basis. Thus, the relationship between ESAL and skid resistance might not be precisely identified. Furthermore, pavement material can be very different in each location depending on the local source of materials. These aspects could influence the results. Therefore, to identify the relationship between ESAL and skid resistance is a future task, and it is recommended that continuous observation of skid resistance be conducted to solve this issue [51][52][53][22]. Finally, from the DT analysis, the methodology of ML is a promising technique that can genuinely capture the variety of patterns from safety road data and overcome uncertainty. Some benefits can be explored and used in the future, such as real-time analysis and decision-maker support. The application of ML opens new doors to gathering more data and applying valuable inputs that can present a wider picture of accidents. The technique leads to automation of the field and allows the process to be smarter. In this case, the importance of gathering more attributes in the future will be essential for the implementation of advanced analysis for reducing SVA [54].

10. Conclusions

MECL has been established to reduce traffic congestion around the Tokyo area, and traffic congestion has significantly decreased. However, road safety on the expressway has been a big issue. Although there has been a decrease in traffic congestion, the number of single-vehicle accidents (SVA) has increased because AVS has risen. Therefore, this study focuses on SVA and aims to identify the conceivable factors causing SVA from the viewpoint of uncertainty, sustainability and quantitative road factors. The identification of specific local parameters that influence the road safety will improve societal sustainability. Weather conditions, skid resistance, AADT and AVS have been defined as uncertainty parameters. Furthermore, curve radius in each location, which is a representative physical condition of the road, has been defined as a quantitative factor. This study collects 10 years of data sets from 2009 to 2019. From the analysis, ML can improve safety, manage risks, analyze and capture the hidden patterns in the data and address accidents. Analyzing road accidents can be performed locally or internationally and presents the root cause of the incidents and the correlations between many factors in different systems accurately. In this study, Pearson's correlation analysis and multiple regression analysis have been adopted. Then, the influence of ESAL, which has not been considered by MECL so far, is determined using the correlation analysis.

- First, in Pearson's correlation analyzes, weather conditions are found to influence SVA significantly. In particular, rainy days, which make the pavements' surface wet, have caused a large number of SVA or about 6.7 times higher than those incurred in the days without precipitation which forms 68% (see Figure 25). In addition, AADT and curve radius could have an effect on SVA in dry pavement conditions. On the other hand, the combination of skid resistance and AVS could also have an influence on SVA in wet conditions. As a result, the significance of each factor might depend on the pavement condition.
- Second, another method for correlation analysis called "cross tabulation analysis" has been applied to further understand the detailed characteristics of the correlation. As a result of the analysis, in the case of less than AADT 30000 vehicles/day in dry pavement conditions, it is clear that even if AVS increases, SVA does not increase. On the other hand, a sharper curve radius could increase the number of accidents in the same situation. In wet pavement conditions, SVA could increase with an increase in AVS. In contrast, SVA could fall with an increase in skid resistance.
- Third, multiple regression analysis is applied based on the correlation results of the cross-tabulation analyses. This analysis aims to identify the significance in terms of the effect on SVA due to the combination of independent variables. Consequently, two statistically meaningful results can be identified. It has become apparent that AVS is an important factor as well as the

skid resistance being a very important factor in wet pavement conditions. Consequently, a suitable management of AVS and skid resistance is considered to be one of the most important strategies for MECL to mitigate SVA.

- Fourth, this study ascertains that there is a clear relationship between skid resistance and ESAL as many other studies identified. It also identifies whether ESAL would be applicable for the safety management of the Metropolitan Expressway. Based on the result, the relationship could not be clearly identified. This might be because there were insufficient ESAL data. However, if monthly data are considered, it can be observed that the skid resistance increases after any pavement reconstruction.
- Fifth, the ML is the future technology which provides advanced analysis comparing with the traditional research (Statistical models) as well as offering precise outcomes and other benefits supporting decision-making on time and ability to learn from the road safety data.

Lastly, considering all of the results, this study could clearly suggest suitable and sustainable countermeasures corresponding to each of the conditions referred to. The countermeasures are summarized in Table 17, where the darker the color of the results implies the more likelihood there is it will occur. When considering the reduction of the number of SVA, accidents in wet pavement conditions have to be analyzed because the wet pavement conditions can cause the frequency of SVA to be seven times higher than those in dry pavement conditions. This implies that the skid resistance needs to be improved, or the facility, which encourages drivers to reduce vehicle speed, needs to be placed in order to reduce accidents. This study has identified the influential causes of SVA and effective countermeasures especially in wet pavement conditions. These findings are now provided to MECL and this policy enables a more sustainable strategy in asset management. Although this study could not identify a clear relationship between skid resistance and ESAL, the apparent relationship between SVA and skid resistance has been demonstrated. For future work, the researchers will be focused on predicting and analyzing accidents on the road using more data, including more road safety external parameters.

Table 14. Sustainable countermeasure policies corresponding to each independent variable condition.

Pavement condition	Condition of each aspect			Results	Countermeasure
	AADT	AVS	Curve radius		
Dry	-30,000 vehicles/day	-	-	SVA will not increase even if AVS increase	・Place sign boards and marks which encourage drivers to be vigilant about the curves
		-	-	SVA will reduce if curving radius increase	
		+60 km/h	-	SVA will reduce if curving radius increase	・Place sign boards and marks which encourage drivers to reduce vehicle speed
	-		-	SVA will increase if AADT increases	
	-		+100 m	SVA will increase if AADT increases	
	-			SVA will increase if AADT increases	・Place sign boards and marks which encourage drivers to be vigilant about the curves
Wet	-	-60 km/h	-	SVA will reduce if skid resistance increases	・Improve skid resistance of a pavement
	-		+100 m	SVA will increase if AVS increase	・Place sign boards and marks which encourage drivers to reduce vehicle speed
	-	-		SVA will increase if AVS increases	・Place sign boards and marks which encourage drivers to reduce vehicle speed
	-	-	-100 m	SVA will reduce if skid resistance increase	・Improve skid resistance of a pavement
	-	-60 km/h		SVA will reduce if skid resistance increase	

Author Contributions: "Conceptualization, T.O., M.S. and S.K.; methodology, T.O., M.S. and S.K.; software, T.O., M.S. and Q.Y.M.; validation, T.O., M.S. and S.K.; formal analysis, T.O.; investigation, T.O., M.S. and S.K.; resources, S.K.; data curation, T.O. and M.S.; writing—original draft preparation, T.O. and M.S.; writing—review and editing, T.O., M.S. and S.K.; visualization, T.O.; supervision, S.K.; project administration, S.K.; funding acquisition, S.K. All authors have read and agreed to the published version of the manuscript."

Funding: This research was funded by European Commission, grant numbers: H2020-MSCA-RISE No. 691135.

Acknowledgments:

The authors wish to thank Senior division manager, Mr. Motoyasu Saito, for his very helpful advice. We are grateful also to Metropolitan Expressway Company Limited for their direction and the data they provided. This work also would not have been possible without the dedicated cooperation of Metropolitan Expressway colleagues. Finally, the authors are sincerely grateful to the European Commission for the financial sponsorship of the H2020-MSCA-RISE Project No. 691135 "RISEN: Rail Infrastructure Systems Engineering Network," which enables a global research network to tackle the grand challenge of railway infrastructure resilience and advanced sensing in extreme environments (www.risen2rail.eu)[55].

Conflicts of Interest: The authors declare no conflict of interest.

Data Availability: All data, models, and code generated or used during the study appear in the submitted article.

References

1. Li, Z.; Liu, P.; Wang, W.; Xu, C. Development of a control strategy of variable speed limits to reduce rear-end collision risks near freeway recurrent bottlenecks. *IEEE Trans. Intell. Transp. Syst.* **2014**, *15*, 866–877, doi:10.1109/TITS.2013.2293199.
2. Metropolitan Expressway Company Limited (MECL) Company Profile 2019-2020 Available online: https://www.shutoko.co.jp/~/Media/pdf/index/about/profile/booklet_profile_2019.pdf (accessed on Feb 19, 2021).
3. Metropolitan Expressway Company Limited (MECL) Driving on the Metropolitan Expressway | shutoko | Available online: https://www.shutoko.co.jp/en/index/driving/howto/ (accessed on Feb 19, 2021).
4. Tanishita, M.; van Wee, B. Impact of vehicle speeds and changes in mean speeds on per vehicle-kilometer traffic accident rates in Japan. *IATSS Res.* **2017**, *41*, 107–112, doi:10.1016/j.iatssr.2016.09.003.
5. Zhu, H.; Dixon, K.K.; Washington, S.; Jared, D.M. Predicting single-vehicle fatal crashes for two-lane rural highways in Southeastern United States. *Transp. Res. Rec.* **2010**, 88–96, doi:10.3141/2147-11.
6. Dong, B.; Ma, X.; Chen, F.; Chen, S. Investigating the differences of single-vehicle and multivehicle accident probability using mixed logit model. *J. Adv. Transp.* **2018**, *2018*, doi:10.1155/2018/2702360.
7. Lee, C.; Li, X. Analysis of injury severity of drivers involved in single- and two-vehicle crashes on highways in Ontario. *Accid. Anal. Prev.* **2014**, *71*, 286–295, doi:10.1016/j.aap.2014.06.008.
8. Hou, G.; Chen, S.; Chen, F. Framework of simulation-based vehicle safety performance assessment of highway system under hazardous driving conditions. *Transp. Res. Part C Emerg. Technol.* **2019**, *105*, 23–36, doi:10.1016/j.trc.2019.05.035.

9. Raff S, M. INTERSTATE HIGHWAY - ACCIDENT STUDY. *Highw. Res. Board Bull.* **1953**, *74*, 18–45.

10. Pardillo Mayora, J.M.; Jurado Piña, R. An assessment of the skid resistance effect on traffic safety under wet-pavement conditions. *Accid. Anal. Prev.* **2009**, *41*, 881–886, doi:10.1016/j.aap.2009.05.004.

11. Islam Bin, M.; Kanitpong, K. Identification of factors in road accidents through in-depth accident analysis. *IATSS Res.* **2008**, *32*, 58–67, doi:10.1016/s0386-1112(14)60209-0.

12. Kaewunruen, S.; Sresakoolchai, J.; Zhou, Z. Sustainability-based lifecycle management for bridge infrastructure using 6D BIM. *Sustain.* **2020**, *12*, 2436, doi:10.3390/su12062436.

13. Lee, J.; Chae, J.; Yoon, T.; Yang, H. Traffic accident severity analysis with rain-related factors using structural equation modeling – A case study of Seoul City. *Accid. Anal. Prev.* **2018**, *112*, 1–10, doi:10.1016/j.aap.2017.12.013.

14. Yau, K.K.W. Risk factors affecting the severity of single vehicle traffic accidents in Hong Kong. *Accid. Anal. Prev.* **2004**, *36*, 333–340, doi:10.1016/S0001-4575(03)00012-5.

15. Sa'adin, S.L.B.; Kaewunruen, S.; Jaroszweski, D. Risks of climate change with respect to the Singapore-Malaysia high speed rail system. *Climate* 2016, *4*.

16. Omura, T.; Kaewunruen, S. Viaduct maintenance for future traffic demands and earthquakes. *Infrastruct. Asset Manag.* **2020**, *7*, 256–268, doi:10.1680/jinam.19.00063.

17. Komada, Y.; Asaoka, S.; Abe, T.; Inoue, Y. Short sleep duration, sleep disorders, and traffic accidents. *IATSS Res.* 2013, *37*, 1–7.

18. Young, M.S. Human Factors of Visual and Cognitive Performance in Driving. *Ergonomics* **2010**, *53*, 444–445, doi:10.1080/00140130903494785.

19. Nishiuchi, H. Traffic Safety Education and Awareness Activities in Japan. **2014**.

20. Nishitani, Y. Alcohol and traffic accidents in Japan. *IATSS Res.* 2019, *43*, 79–83.

21. Hung, K.V.; Huyen, L.T. Education influence in traffic safety: A case study in Vietnam. *IATSS Res.* **2011**, *34*, 87–93, doi:10.1016/j.iatssr.2011.01.004.

22. Rolison, J.J.; Regev, S.; Moutari, S.; Feeney, A. What are the factors that contribute to road accidents? An assessment of law enforcement views, ordinary drivers' opinions, and road accident records. *Accid. Anal. Prev.* **2018**, *115*, 11–24, doi:10.1016/j.aap.2018.02.025.

23. Papadimitriou, E.; Filtness, A.; Theofilatos, A.; Ziakopoulos, A.; Quigley, C.; Yannis, G. Review and ranking of crash risk factors related to the road infrastructure. *Accid. Anal. Prev.* **2019**, *125*, 85–97, doi:10.1016/j.aap.2019.01.002.

24. Oguchi, T. Achieving safe road traffic - the experience in Japan. *IATSS Res.* **2016**, *39*, 110–116, doi:10.1016/j.iatssr.2016.01.003.

25. Uchida, N.; Kawakoshi, M.; Tagawa, T.; Mochida, T. An investigation of factors contributing to major crash types in Japan based on naturalistic driving data. *IATSS Res.* **2010**, *34*, 22–30, doi:10.1016/j.iatssr.2010.07.002.

26. Buss, D.; Abishev, K.; Baltabekova, A. Driver's reliability and its effect on road traffic safety. In Proceedings of the Procedia Computer Science; Elsevier B.V., 2019; Vol. 149, pp. 463–466.

27. Jung, S.; Jang, K.; Yoon, Y.; Kang, S. Contributing factors to vehicle to vehicle crash frequency and severity under rainfall. *J. Safety Res.* **2014**, *50*, 1–10, doi:10.1016/j.jsr.2014.01.001.

28. Malin, F.; Norros, I.; Innamaa, S. Accident risk of road and weather conditions on different road types. *Accid. Anal. Prev.* **2019**, *122*, 181–188, doi:10.1016/j.aap.2018.10.014.

29. Üzümcüoğlu, Y.; Özkan, T.; Wu, C.; Zhang, H. How drivers perceive traffic? How they behave in traffic of Turkey and China? *Transp. Res. Part F Traffic Psychol. Behav.* **2019**, *64*, 463–471, doi:10.1016/j.trf.2019.06.006.

30. Dadashova, B.; Ramírez, B.A.; McWilliams, J.M.; Izquierdo, F.A. The Identification of Patterns of Interurban Road Accident Frequency and Severity Using Road Geometry and Traffic Indicators. In Proceedings of the Transportation Research Procedia; Elsevier B.V., 2016; Vol. 14, pp. 4122–4129.

31. Yan, X.; Radwan, E.; Abdel-Aty, M. Characteristics of rear-end accidents at signalized intersections using multiple logistic regression model. *Accid. Anal. Prev.* **2005**, *37*, 983–995, doi:10.1016/j.aap.2005.05.001.

32. Shong, N. Pearson's versus Spearman's and Kendall's correlation coefficients for continuous data, 2010.

33. Smalheiser, N.R. *Data literacy: How to make your experiments robust and reproducible*; 2017; ISBN 9780128113066.

34. Theofilatos, A.; Yannis, G. A review of the effect of traffic and weather characteristics on road safety. *Accid. Anal. Prev.* **2014**, *72*, 244–256, doi:10.1016/j.aap.2014.06.017.

35. Chen, C.; Xie, Y. Modeling the effects of AADT on predicting multiple-vehicle crashes at urban and suburban signalized intersections. *Accid. Anal. Prev.* **2016**, *91*, 72–83, doi:10.1016/j.aap.2016.02.016.

36. Cohen, P. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*; 2014;

37. Borowitz, M. Japan Meteorological Agency. In *Open Space*; 2018.

38. Mukaka, M.M. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi Med. J.* **2012**, *24*, 69–71.

39. Nau, R. What ' s a good value for R-squared ? Available online: https://people.duke.edu/~rnau/rsquared.htm (accessed on Feb 26, 2021).

40. Ezra, H. On the relationship between road safety research and the practice of road design and operation. *Accid. Anal. Prev.* **2019**, *128*, 114–131, doi:10.1016/j.aap.2019.03.016.

41. Boujemaa, K.S.; Berrada, I.; Fardousse, K.; Naggar, O.; Bourzeix, F. Toward Road Safety Recommender Systems: Formal Concepts and Technical Basics. *IEEE Trans. Intell. Transp. Syst.* **2021**, doi:10.1109/TITS.2021.3052771.

42. Jain, A.; Ahuja, G.; Anuranjana; Mehrotra, D. Data mining approach to analyse the road accidents in India. In Proceedings of the 2016 5th International Conference on Reliability, Infocom Technologies and Optimization, ICRITO 2016: Trends and Future Directions; Institute of Electrical and Electronics Engineers Inc., 2016; pp. 175–179.

43. Mannering, F.; Bhat, C.R.; Shankar, V.; Abdel-Aty, M. Big data, traditional data and the tradeoffs between prediction and causality in highway-safety analysis. *Anal. Methods Accid. Res.* **2020**, *25*, 100113, doi:10.1016/j.amar.2020.100113.

44. Li, L.; Shrestha, S.; Hu, G. Analysis of road traffic fatal accidents using data mining techniques. In Proceedings of the Proceedings - 2017 15th IEEE/ACIS International Conference on Software Engineering Research, Management and Applications, SERA 2017; 2017; pp. 363–

370.

45. Lian, Y.; Zhang, G.; Lee, J.; Huang, H. Review on big data applications in safety research of intelligent transportation systems and connected/automated vehicles. *Accid. Anal. Prev.* **2020**, *146*, doi:10.1016/j.aap.2020.105711.

46. Mahajan, N.; Pal, B. Analysis of Factors of Road Traffic Accidents using Enhanced Decision Tree Algorithm. *Int. J. Comput. Appl.* **2016**, *135*, 1–3, doi:10.5120/ijca2016908231.

47. Aljoufie, M.; Tiwari, A. Modeling Road Safety in Car-Dependent Cities: Case of Jeddah City, Saudi Arabia. *Sustainability* **2021**, *13*, 1816, doi:10.3390/su13041816.

48. Ghandour, A.J.; Hammoud, H.; Al-Hajj, S. Analyzing factors associated with fatal road crashes: A machine learning approach. *Int. J. Environ. Res. Public Health* **2020**, *17*, 4111, doi:10.3390/ijerph17114111.

49. Alawad, H.; Kaewunruen, S. Wireless Sensor Networks: Toward Smarter Railway Stations. *Infrastructures* **2018**, *3*, 24, doi:10.3390/infrastructures3030024.

50. Alawad, H.; Kaewunruen, S. 5G Intelligence Underpinning Railway Safety in the COVID-19 Era. *Front. Built Environ.* **2021**, *7*, 14, doi:10.3389/fbuil.2021.639753.

51. TRL Overseas Road Note 18 - A guide to the pavement evaluation and maintenance of bitumen-surfaced roads in tropical and sub-tropical countries. *Overseas road note 31 (ORN 31), 4th Ed., Transp. Road Res. Lab. Crowthorne, London, UK* **1999**, 1–83.

52. Subirats, P.; Goyat, Y.; Jacob, B.; Violette, E. A New Road Safety Indicator Based on Vehicle Trajectory Analysis. In Proceedings of the Transportation Research Procedia; Elsevier B.V., 2016; Vol. 14, pp. 4267–4276.

53. Tsubota, T.; Fernando, C.; Yoshii, T.; Shirayanagi, H. Effect of Road Pavement Types and Ages on Traffic Accident Risks. In Proceedings of the Transportation Research Procedia; Elsevier B.V., 2018; Vol. 34, pp. 211–218.

54. Alawad, H.; Kaewunruen, S.; An, M. Learning from Accidents: Machine Learning for Safety at Railway Stations. *IEEE Access* **2020**, *8*, 633–648, doi:10.1109/ACCESS.2019.2962072.

55. Kaewunruen, S.; Sussman, J.M.; Matsumoto, A. Grand challenges in transportation and transit systems. *Front. Built Environ.* **2016**, *2*, 1–5, doi:10.3389/fbuil.2016.00004.