공학박사 학위논문

The Effect of Section Speed Enforcement System on Casualty Crash Reduction using Interpretable Machine Learning

Interpretable Machine Learning을 활용한 구간단속시스템 설치에 따른 인명피해사고 감소 효과 연구

2020년 8월

서울대학교 대학원

공과대학 건설환경공학부

홍 경 식

The Effect of Section Speed Enforcement System on Casualty Crash Reduction using Interpretable Machine Learning

지도교수 김 동 규

이 논문을 공학박사 학위논문으로 제출함

2020년 5월

서울대학교 대학원

공과대학 건설환경공학부

홍 경 식

홍경식의 박사 학위논문을 인준함

2020년 7월

위 원 장	고 승 영	(인)
부위원장	김 동 규	(인)
위 원	이 청 원	(인)
위 원	박 신 형	(인)
위 원	장 기 태	(인)

Abstract

In this study, a prediction model for casualty crash occurrence was developed considering whether to install SSES and the effect of SSES installation was quantified by dividing it into direct and indirect effects through the analysis of mediation effect. Also, it was recommended what needs to be considered in selecting the candidate sites for SSES installation. For this, crash prediction model was developed by using the machine learning for binary classification based on whether or not casualty crash occurred and the effects of SSES installation were analyzed based on crashes and speed-related variables. Especially, the IML methodology was applied that considered the predictive performance as well as the interpretability of the forecast results as important. When developing the IML which consisted of black-box and interpretable model, KNN, RF, and SVM were reviewed as black-box model, and DT and BLR were reviewed as interpretable model. In the model development, the hyper-parameters that could be set in each methodology were optimized through k-fold cross validation. The SVM with a polynomial kernel trick was selected as black-box model and the BLR was selected as interpretable model to predict the probability of casualty crash occurrence.

For the developed IML model, the evaluation was conducted through comparison with the typical BLR from the perspective of the PDR framework. The evaluation confirmed that the results of the IML were more excellent than the typical BLR in terms of predictive accuracy, descriptive accuracy, and relevancy from a human in the loop.

Using the result of IML's model development, the effect on SSES installation were quantified based on the probability equation of casualty crash occurrence. The equation is the logistic function that consists of SSES, SOR, SV, TVL, HVR, and CR. The result of analysis confirmed that the SSES installation reduced the probability of casualty crash occurrence by about 28%. In addition, the analysis of mediation effects on the variables affected by installing SSES was conducted to quantify the direct and indirect effects on the probability of reducing the casualty crashes caused by the SSES installation. The proportion of indirect effects through reducing the ratio of exceeding the speed limit (SOR) was about 30% and the proportion of indirect effects through reduction of speed variance (SV) was not statistically significant at the 95% confidence level.

Finally, the probability equation of casualty crash occurrence developed in this study was applied to the sections of Yeongdong Expressway to compare the crash risk section with the actual crash data to examine the applicability of the development model. The analysis result verified that the equation was reasonable. Therefore, it may be considered to select dangerous sites based on casualty crash and speeding firstly, and then to install SSES at the section where traffic volume (TVL), heavy vehicle ratio (HVR), and curve ratio (CR) are higher than the other sections.

Keywords: binary classification, casualty crash prediction, Interpretable Machine Learning (IML), mediation effect analysis, Section Speed Enforcement System (SSES)

Student Number: 2010-31011

Contents

1. Introduction ··· 1
  1.1. Background of research ··· 1
  1.2. Objective of research ··· 4
  1.3. Research Flow ··· 6

2. Literature Review ··· 11
  2.1. Research related to SSES ··· 11
    2.1.1. Effectiveness of SSES ··· 11
    2.1.2. Installation criteria of SSES ··· 15
  2.2. Machine learning about transportation ··· 17
    2.2.1. Machine learning algorithm ··· 17
    2.2.2. Machine learning algorithm about transportation ··· 19
  2.3. Crash prediction model ··· 23
    2.3.1. Frequency of crashes ··· 23
    2.3.2. Severity of crash ··· 26
  2.4. Interpretable Machine Learning (IML) ··· 31
    2.4.1. Introduction ··· 31
    2.4.2. Application of IML ··· 33

3. Model Specification ······ 37
3.1. Analysis of SSES effectiveness ······ 37
3.1.1. Crashes analysis ······ 37
3.1.2. Speed analysis ······ 39
3.2. Data collection & pre-analysis ······ 40
3.2.1. Data collection ······ 40
3.2.2. Basic statistics of variables ······ 42
3.3. Response variable selection ······ 50
3.4. Model selection ······ 52
3.4.1. Binary classification ······ 52
3.4.2. Accuracy vs. Interpretability ······ 53
3.4.3. Overview of IML ······ 54
3.4.4. Process of model specification ······ 57

4. Model development ······ 59
4.1. Black-box and interpretable model ······ 59
4.1.1. Consists of IML ······ 59
4.1.2. Black-box model ······ 60
4.1.3. Interpretable model ······ 68
4.2. Model development ······ 72
4.2.1. Procedure ······ 72
4.2.2. Measures of effectiveness ······ 74
4.2.3. K-fold cross validation ······ 76

4.3. Result of model development ··· 78
4.3.1. Result of black-box model ··· 78
4.3.2. Result of interpretable model ··· 85

5. Evaluation & Application ··· 91
5.1. Evaluation ··· 91
5.1.1. The PDR framework for IML ··· 91
5.1.2. Predictive accuracy ··· 93
5.1.3. Descriptive accuracy ··· 94
5.1.4. Relevancy ··· 99
5.2. Impact of Casualty Crash Reduction ··· 102
5.2.1. Quantification of the effectiveness ··· 102
5.2.2. Mediation effect analysis ··· 106
5.3. Application for the Korean expressway ··· 118

6. Conclusion ··· 121
6.1. Summary and Findings ··· 121
6.2. Further Research ··· 125

List of Tables

<Table 1-1> Automated traffic enforcement camera system in Korea ··········· 2
<Table 2-1> Effectiveness of SSES ··· 11
<Table 2-2> Installation criteria of SSES ··· 15
<Table 2-3> Summary of reviews on machine learning about transportation ······ 19
<Table 2-4> Summary of reviews on crash frequency ··· 23
<Table 2-5> Summary of reviews on crash severity ··· 26
<Table 2-6> Introduction on IML techniques ··· 31
<Table 2-7> Summary of reviews on IML ··· 33
<Table 3-1> Result of crash analysis (naive before-after test) ··· 37
<Table 3-2> Result of crash analysis (C-G method) ··· 38
<Table 3-3> Result of speed analysis (naive before-after test) ··· 39
<Table 3-4> Result of speed analysis (C-G method) ··· 40
<Table 3-5> Scope of data collection ··· 41
<Table 3-6> Contents of data collection ··· 42
<Table 3-7> Contingency table between SSES and casualty ··· 42
<Table 3-8> Variable description ··· 43
<Table 3-9> Basic statistics value ··· 44
<Table 3-10> Correlation coefficient between variables ··· 46
<Table 3-11> t - test results for casualty ··· 49
<Table 3-12> t - test results for SSES ··· 49
<Table 3-13> Summary of interpretable models classification ··· 57
<Table 4-1> Confusion matrix for MOE ··· 74
<Table 4-2> MOE for machine learning ··· 75
<Table 4-3> Predicted result of KNN ··· 79

<Table 4-4> Predicted result of RF ··· 81
<Table 4-5> Formula and parameters for kernel functions in the SVM ··· 83
<Table 4-6> Best parameter for kernel functions in the SVM ··· 83
<Table 4-7> Predicted result of SVM ··· 84
<Table 4-8> Predicted result of the black-box models ··· 85
<Table 4-9> Result of DT ··· 87
<Table 4-10> Result of BLR ··· 88
<Table 4-11> Predicted result of the interpretable models ··· 88
<Table 5-1> Predictive accuracy ··· 93
<Table 5-2> Descriptive accuracy ··· 94
<Table 5-3> Confusion matrix of IML-BLR ··· 95
<Table 5-4> List of miss-classification ··· 95
<Table 5-5> Result of IML-BLR ··· 100
<Table 5-6> Result of BLR ··· 101
<Table 5-7> Effectiveness of crash reduction in literature reviews ··· 105
<Table 5-8> Type of models possible to estimate mediation effects ··· 111
<Table 5-9> Mediation effect analysis of SOR ··· 113
<Table 5-10> Mediation analysis of SV ··· 117
<Table 5-11> Top 10 sections for probability of casualty crash occurrence ··· 118
<Table 5-12> Top 5 sections for frequency of casualty crashes ··· 119

List of Figures

[Figure 1-1] Configuration of SSES ········ 1

[Figure 1-2] Flowchart of research ········ 7

[Figure 3-1] Scatter plot between variables ········ 45

[Figure 3-2] Box and whisker plot ········ 47

[Figure 3-3] Response variable selection process ········ 50

[Figure 3-4] Relation between interpretability and accuracy ········ 53

[Figure 3-5] Interpretable machine learning ········ 56

[Figure 3-6] Process of model specification ········ 58

[Figure 4-1] Model selection for IML ········ 59

[Figure 4-2] Concept of KNN classification ········ 61

[Figure 4-3] Euclidean distance ········ 62

[Figure 4-4] Working of RF algorithm ········ 65

[Figure 4-5] Working of BLR algorithm ········ 71

[Figure 4-6] Procedure of model development ········ 73

[Figure 4-7] K-fold cross validation ········ 77

[Figure 4-8] Validation for optimal k ········ 78

[Figure 4-9] Tuning the hyper-parameter (m-try) ········ 80

[Figure 4-10] Tuning the hyper-parameter (max-nodes) ········ 80

[Figure 4-11] Tuning the hyper-parameter (n-tree) ········ 81

[Figure 4-12] Black-box model selection ········ 85

[Figure 4-13] Tuning the size of tree ········ 86

[Figure 4-14] Result of DT ········ 86

[Figure 4-15] Tuning the optimal cut-off value ········ 87

[Figure 4-16] Result of model development for IML ········ 89

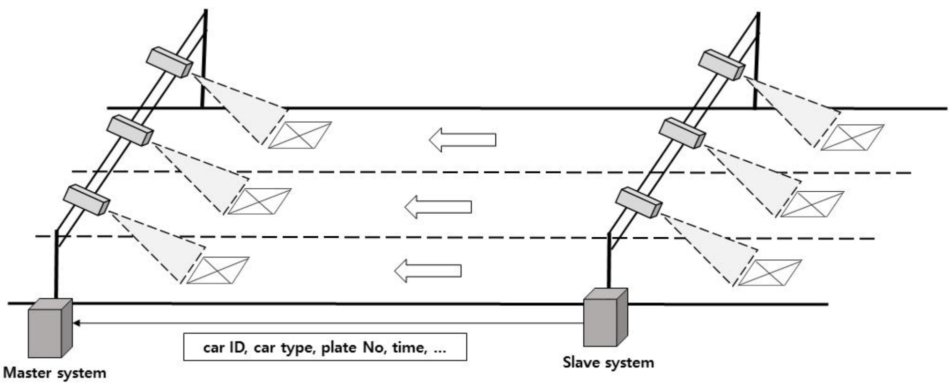
[Figure 5-1] PDR framework for IML ··· 91
[Figure 5-2] ROC and AUC ··· 96
[Figure 5-3] AUROC curve of BLR ··· 98
[Figure 5-4] AUROC curve of IML-BLR ··· 99
[Figure 5-5] Binary logistic function ··· 102
[Figure 5-6] Total effect model ··· 108
[Figure 5-7] Mediation effect model ··· 108
[Figure 5-8] Structure of the mediation package ··· 110
[Figure 5-9] Mediation effect model of SOR ··· 112
[Figure 5-10] sensitivity analysis of SOR ··· 115
[Figure 5-11] Mediation effect model of SV ··· 116

1. Introduction

1.1. Background of research

Because speeding is one of the most significant contributing factors to fatal crashes, most road traffic agencies attempt to achieve the right operating speed by imposing speed limits. Speed limit violations are prevalent, even on roadways with speed cameras. But, a problem with automated speed enforcement system is that some drivers brake before passing a camera location and then exceed the speed limit after passing. This sudden braking can cause dangerous situations, crashes, and traffic jams (Montella 2012).

A new technique to overcome these problems is the Section Speed Enforcement System (SSES). Unlike conventional automated speed enforcement, which measure the speed of a vehicle at one spot, the SSES calculates the average speed over a long distance (at least 500m and up to several kilometers).



[Figure 1-1] Configuration of SSES

SSES has cameras installed on all lanes, slave camera enforcement system is installed at the start point, and master camera enforcement system is installed at the end point as shown in [Figure 1-1]. The operating principles of SSES are as follows. The vehicle passes on the level of the camera which records the number plate and the specific time of passage in the salve camera enforcement system and they are sent to the master camera enforcement system. it passes in front of a second camera which again read the number plate and specific time in the master camera enforcement system. The controller of master system calculates the average speed.

Since its first introduction in the Netherlands, it has been in operation in France, Austria, Germany, UK, Italia, Norway, Australia, New Zealand, etc. In South Korea, the automated traffic enforcement system was introduced at 32 locations nationwide in 1997. The types of automated traffic enforcement system which is installed and operated by Korean National Police Agency (KNPA) are (spot) speed, red-light, SSES and mobile cameras. They have been expanded and installed in succession because the effectiveness of reducing crashes is high. Current state of installation in the Korea is shown in <Table 1-1>.

<Table 1-1> Automated traffic enforcement camera system in Korea

Type	Speed	Red-light	SSES	Mobile	Total
No. of cameras	3,091	5,042	469	399	9,001
Proportion (%)	34.4	56.0	5.2	4.4	100.0

Source: KNPA(2020. 01.)

In case of SSES, a total of 469 cameras have been operating in 97 sections, since it was first set up on the Seohaean expressway in 2007. This accounts for about 5.2% of all automatic traffic enforcement systems. In this regard, the KNPA also acknowledges the need to expand the SSES, which has a greater effectiveness of preventing casualty crashes and stabilizing traffic flow compared with other automated traffic enforcement systems. However, it is difficult to expand the installation of SSES because there is no quantitative installation criteria.

When reviewing research related to SSES, most studies on installation effectiveness are focused on speed, crash, and environmental pollutant emissions. The effectiveness analysis for SSES installation is being performed using naive before-after test, Comparison-Group (C-G) method, and Empirical Bayes (EB) method. In addition, some studies have been conducted on the installation criteria of SSES, mostly in the form of suggestions for qualitative criteria rather than quantitative ones. The qualitative criteria suggest that crash frequency, crash severity, speed, proportion of exceeding speed limit, traffic volume, and heavy vehicle ratio should be considered when selecting the location for the installation of SSES.

When reviewing an crash prediction model related to SSES, it is mainly a model that predicts crash frequency or crash severity. Most of the studies, the prediction for crash frequency is developed by applying Generalized Linear Model (GLM) such as Poisson model and negative binomial model, and the prediction for crash severity is developed by applying machine learning using classification techniques such as Binary Logistic Regression (BLR), Random Forest (RF), Support Vector Machine (SVM) and Artificial Neural Network (ANN). There is no model that considered SSES as a independent variable when developing the crash prediction model. Therefore, it is necessary to develop an crash prediction model to quantify the installation effectiveness of SSES and to make suggestions on what needs to be considered in selecting the location for SSES installation.

1.2. Objective of research

The purpose of this study is to develop the prediction model of casualty crash occurrence, to quantify the effectiveness of SSES installation and to make suggestions on what needs to be considered in selecting the location for SSES installation. To achieve the purpose of the study, it is important to improve the prediction accuracy for the prediction model of casualty crash occurrence. In addition, the interpretability of prediction model is also important to quantify the effect of SSES on casualty crash reduction and to recommend the candidate sites for SSES installation. Therefore, Interpretable Machine Learning (IML) methodology is applied to improve the model's performance and interpretability in the model development. IML is a methodology that has been introduced to increase the ability to explain machine learning techniques that have excellent predictive performance, such as RF, SVM, and DNN, but lack the ability to interpret forecast results, and has been actively researched in medicine and engineering.

To quantify the installation effects of SSES, a model for probability of casualty crash occurrence is developed and the indirect effects of variables (e.g. mean speed, the ratio of exceeding the speed limit and speed variance) related to speed are analyzed separately from the direct effects of reducing probability of casualty crash occurrence caused by installation of SSES. For this, the process of mediation effect analysis is carried out. And the methodology is proposed to select candidates for installation of SSES based on the developed probability formula of casualty crash occurrence.

The differentiations in this study from prior studies are as follows. First of all, there is no crash prediction model considering whether or not SSES is installed. It is necessary to develop an crash prediction model to quantify the installation effects of SSES.

Secondly, many prior studies have analyzed the effectiveness of crash reduction before-after SSES installation, but this study quantifies the effects of SSES installation by developing a prediction model for the probability of casualty crash occurrence. In other word, the effects of reducing the number of crashes were analyzed in prior studies, but in this study, the effects of reducing probability of casualty crash occurrence are analyzed.

Thirdly, when the effectiveness of SSES installation is quantified, direct effects on the reduction of casualty crash caused by the installation of SSES and indirect effects by the induction of speed reduction are divided through the analysis of the mediation effects.

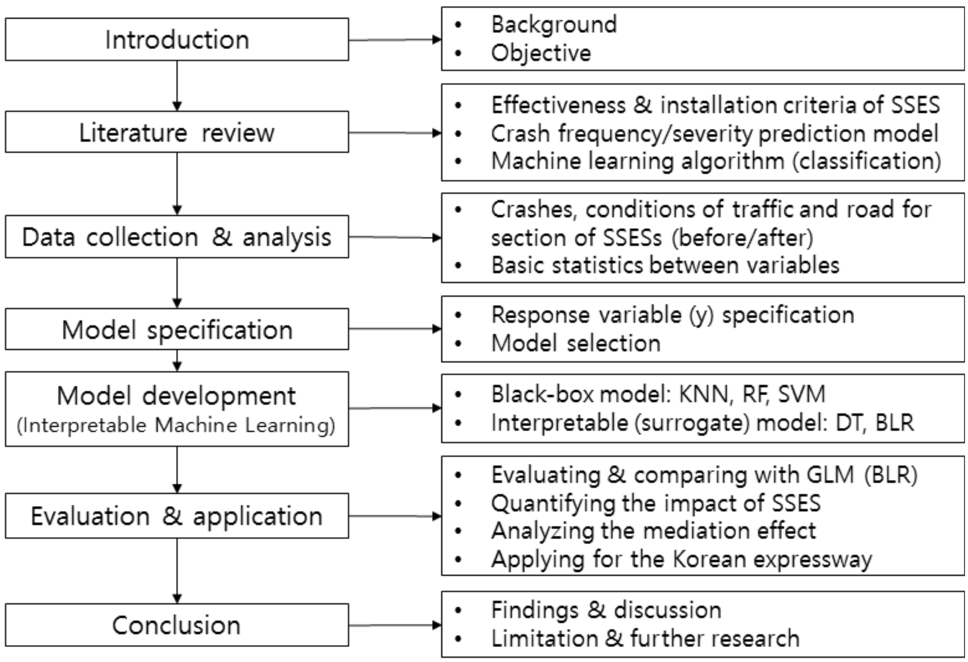
Finally, in this study, it is recommended what needs to be considered in selecting the location for SSES installation based on the result of developing crash prediction models. Through this, it can support the decision making of KNPA when installing SSES, and can also be used as a basic qualitative criteria to select candidate locations for SSES installation.

1.3. Research Flow

The purpose of this study is to develop the prediction model of casualty crash occurrence using the IML and to quantify the effectiveness of SSES installation through mediation effect analysis. In order to achieve the purpose of study, the following processes and methodologies have been carried out. The overall research flow of this study is shown in [Figure 1-2] and each chapter covers the following contents.

In the chapter 2, literature review is conducted to set the direction of model development. the prior studies on the effectiveness of SSES, the installation criteria of SSES, crash prediction model and machine learning algorithm are reviewed. First of all, in the research on the analysis of SSES effects, the methodology of analyzing the effect and analysis results are reviewed. Secondly, the installation criteria of SSES in foreign countries are reviewed and compared with the domestic criteria. Thirdly, the crash prediction model are reviewed.

Crash prediction models are largely divided into those that predict the frequency and severity of crashes. Through a review of the model, implications and differentiations for the model to apply in this study are reviewed. Finally, researches on machine learning algorithms for binary classification are reviewed (e.g. parametric methodologies such as DT and BLR, and non-parametric methodologies such as KNN, RF, SVM, and DNN). Especially, researches on the IML for definition, method and application are also reviewed.



[Figure 1-2] Flowchart of research

In the chapter 3, the process of model specification has been carried out. For its purpose, data collection is conducted on SSES locations installed in the Korean expressway. Road, traffic and control conditions which are used as independent variables are collected. Basic statistics such as scatter plot, correlation and box plot between variables are analyzed for data's refining and filtering. The effect analysis of SSES installation is conducted with the Measures Of Effectiveness (MOE) such as speeds and crashes. A specification of the response variable applicable to the model development is carried out. Next, the applicability of the IML techniques for the development of the casualty crash model are reviewed. In addition, the methodologies for mediation effect analysis are reviewed to quantify the effects of SSES installation separately from the direct and indirect effects.

In the chapter 4, prediction model for casualty crash occurrence is developed considering whether or not the SSES installation. The developed prediction model is applied with machine learning for binary classification to predict whether or not an casualty crash has occurred. IML is used to improve the prediction model's performance and interpretability. It usually uses non-parametric method as black-box model for improving the accuracy of prediction and parametric method as interpretable (surrogate) model for improving the interpretability. In this study, KNN, RF, and SVM are applied to black-box models, and Decision Tree (DT) and BLR are applied to interpretable models.

In the chapter 5, a performance evaluation is conducted against the developed IML model compared with the typical BLR model in the perspective of the PDR (Predictive accuracy, Descriptive accuracy and Relevancy) framework. Based on the IML model developed, the effects of casualty crash reduction due to SSES installation are quantified, and the effects of SSES installation are analyzed by separating it by direct and indirect effects through the analysis of mediation effects. Finally, it is suggested what needs to be considered in selecting the location for SSES installation based on the probability formula for casualty crash occurrence.

In the final chapter 6, the findings of this study are summarized and it is reviewed that they can be used to implement policies by the KNPA related to the installation and operation of SSES. Finally, the limitations of the study results are reviewed and the directions of future research are suggested.

2. Literature Review

2.1. Research related to SSES

2.1.1. Effectiveness of SSES

In most prior studies, MOEs are used in terms of traffic safety, operation and environment when analyzing the effects of installing the SSES as shown in <Table 2-1>.

<Table 2-1> Effectiveness of SSES

Author	Year	Subject	Methodology	Results
Torre et al.	2019	Safety effects of automated section speed control on the Italian motorway network	EB analysis	PDO crash: 22%↓ Fatal injury 18%↓
Montella et al.	2015	Effect on speed and safety of point-to-point speed enforcement systems	EB analysis	Stdev of speed: 26%↓ % of exceeding speed limit: 77~84%↓ Total crash: 22%↓
Cascetta et al.	2011	Effects of section speed enforcement system on traffic flow at freeway bottlenecks	Empirical analysis	Mean speed↓ speed variation↓ Bottleneck↓
Jung et al.	2014	Traffic accident reduction effects of section speed enforcement system (SSES) operation in freeways	C-G method	Total crash: 32% ↓ Fatal injury 42%↓
Yun	2011	Effect of the point-to-point speed enforcement system	C-G method	Total crash: 50%↓
Thornton	2010	Reduction in CO_2 emissions and fuel consumption with SSES	Empirical analysis	CO_2 emission: 11%↓ fuel consumption: 30%↓

Torre et al. (2019) evaluated the impact of the Automated Section Speed Control (ASSC) system on the expected crash frequency using

Empirical Bayes (EB) methodology. This study was carried out on a sample of 125 ASSC sites of the Italian motorway network covering 1,252km, where a total of 21,721 crashes were recorded during a 10-year analysis period from 2004 to 2013. The EB analysis estimated a significant 22% reduction in the expected crash frequency due to the implementation of the ASSC system. The analysis indicated that the effect is slightly larger on Property Damage Only (PDO) crashes (−23%) than on fatal injury (FI) crashes (−18%), and that the highest reductions in crash frequency are expected for multi-vehicle FI crashes (−25%) and multi-vehicle PDO crashes (−31%). Furthermore, the results indicated that the ASSC system was more effective in reducing crash rates when traffic volume increased and it was therefore strongly recommended as a countermeasure to improve safety on high traffic volume motorway sections.

Montella et al. (2015) evaluated the effects on speed and safety of the point-to-point (P2P) speed enforcement system activated on the urban motorway A56 in Italy. The P2P system led to very positive effects on both speed and safety. As far as the effects on the section average travel speeds, the system yielded to a reduction in the mean speed, the 85th percentile speed, the standard deviation of speed, and the proportion of exceeding the speed limits, exceeding the speed limits more than 10km/h, and exceeding the speed limits more than 20km/h. The best results were the decrease of the speed variability and the reduction of the excessive speeding behaviour. The decrease in the standard deviation of speed was 26% while the proportion of light and heavy vehicles exceeding the speed limits more than 20km/h was reduced respectively by 84 and 77%. As far as the safety effects, the P2P system yielded to a 32% reduction in the total crashes, with a lower 95% confidence limit of the estimate equal to 22%. The greatest crash reductions were in rainy weather (57%), on wet pavement (51%), on curves (49%), for single vehicle crashes (44%), and for injury crashes (37%).

Cascetta (2011) analyzed the traffic flow conditions (bottleneck phenomenon) before and after the installation of the section control equipment using an empirical analysis. Gathered data consisted of point measurements at detectors and average travel speeds of each vehicle crossing the stretch. The main observed features were following;

- a strong homogenization of individual speeds and of mean speeds among the lanes,
- a reduction in the strength of the bottleneck,
- the emergence of significant oscillations in time of traffic characteristics,
- a sensible reduction of travel times during the congestion pattern caused by the bottleneck moving down-stream of the section.

Empirical evidence suggested that driver compliance with speed limits was the key factor in analysis of such speed management systems and that their concurrent application with dynamic speed limit strategies should be thoroughly evaluated with a particular focus on this measure.

Jeong et al. (2014) analyzed the effects of crash reduction by using the C-G method for SSES operation sections. The number of crashes was reduced by 32.0%, the number of casualty crashes was reduced by 17.1%, and the number of fatal crashes was reduced by 41.7%.

Yun (2011) conducted an analysis of the installation effect of the SSES using the C-G method. The scope of study was analyzed for the number of crashes during one year in three SSES sections installed in 2008, and the result of C-G method showed that the crash reduction was by 49.97%.

Thornton (2010) analyzed annual reduction of passenger car CO_2 emissions and fuel consumption. The analysis results showed that the fuel consumption and CO_2 emissions of passenger cars were reduced by 11% by installing the SSES on highways, and that the speed limit of 50 mph could be reduced by up to 30% by the fuel consumption and CO_2 emissions. The greater the variation in the speed of traffic, the more frequent the braking conditions of the vehicles occurred, and the subsequent driver of the preceding vehicle would also apply the brakes, which in turn caused increased fuel consumption and CO_2 emissions, so smooth driving through the implementation of sectional speeding had been identified as reducing fuel consumption and CO_2 emissions. Further, it was analyzed that the effect of reducing congestion during peak hours was demonstrated in the section of the road construction due to the implementation of section speeding, and that it had a positive effect on reducing the fuel consumption and traffic congestion.

In prior studies, the analysis results showed that installation of the SSES was very effective in reducing the number of crashes and the crash severity, and it also affected the reduction of the vehicle's driving speed and ratio of exceeding the speed limit. In addition, by uniformizing traffic flows of vehicles through install the SSES, the incidental effects of reducing pollutant emissions such as CO_2 were identified.

2.1.2. Installation criteria of SSES

In order to recommend the installation criteria for SSES, one of the purposes of this study, the prior researches related to the installation criteria are reviewed. SSES is widely installed and being operated in the U.S., Australia and Europe, and the installation criteria presented in major countries are shown in <Table 2-2>.

<Table 2-2> Installation criteria of SSES

Nation	Installation criteria (accident, speed, etc.)	AADT	Length
Australia (NSW)	• Crash frequency of the heavy vehicle • Proportion of exceeding the speed limit for the heavy vehicle	-	6~75km
New Zealand	• No. of crashes, crash severity • Without sections in which enforcement is avoided	> 15,000	Over 2km
UK	• Installation criteria of spot enforcement: 3 KSI/Km * KSI (Killed or Seriously Injured) • More than 3 spots in the section of SSES	-	5~20km
Norway	• Mean speed>Speed limit • Same speed limit for the entire section	Not any exit > 250	2~10km

First of all, Australia's NSW state is operating SSES for the purpose of reducing crashes to heavy vehicles. Therefore, it is recommended to install the SSES based on the crash data related to the heavy vehicle and the rate of exceeding speed limit for the heavy vehicle. In addition, the section length of SSES is recommended to set in the range of approximately 6 to 75km.

In New Zealand, the installation site of SSES is selected based on the crash frequency and the crash severity. It is recommended to avoid sections which contain intersections. Also, It is required that daily average traffic volume is more than 15,000 vehicles in the section, and section length is more than 2km (Lynch 2011).

In the U.K., if there are more than three KSI (Killed or Seriously Injured) crashes per km annually, it is required to be selected as candidate site for installing the spot speed enforcement system. And if there are more than three candidate sites of the spot speed enforcement within a given section, the SSES should be installed. The length of the section is to be set in the range of 5 to 10km (DfT 2007).

In the Norway, SSES should be installed in sections where the average speed exceeds the speed limit and the speed limit remains the same for the entire section. It is required that traffic volume to diverge or to merge within the control section is less than 250 per day. In addition, the section length of SSES is recommended to set in the range of about 2 to 10 km (Ragnøy 2011).

As discussed above, the installation criteria for SSES, including the number of crashes, crash severity, speed, heavy vehicle ratio, and uninterrupted traffic flow sections are provided. These installation criteria are qualitative rather than quantitative.

In Korea, there are no specific criteria for installation provided by the KNPA, and each local police agency that is responsible for the installation of SSES. Generally, They select candidate sites considering the number of crashes, Equivalent Property Damage Only (EPDO) and feasibility of installing SSES at the sites.

2.2. Machine learning about transportation

2.2.1. Machine learning algorithm

Arthur Samuel defined machine learning as "a field of research that allows computers to learn without explicitly being programmed". There are three kinds of machine learning: supervised, non-supervised and reinforcement learning.

Supervised learning should include the desired answer or label in the training data that is injected into the algorithm. Classification is a typical map learning task, and number recognition is a good example. Another action is to use a feature called a prediction variable to predict the final result. These kinds of actions are called regression. Some regression algorithms can be used for classification, and sometimes they can't be used. Logistic regression, which is widely used in classification, outputs a probability of belonging to the class.

Below are some of the most important mapping algorithms.

- K-Nearest Neighbors (KNN)

- Binary Logistic Regression (BLR)
- Support Vector Machine (SVM)
- Decision Tree (DT)
- Random Forest (RF)
- Neural Network (NN)

Non-supervised learning does not require the label required for supervised learning. The system must learn without any help.

Below is the most important non-map learning algorithm.

○ Cluster

- K-Means
- Hierarchical Cluster Analysis (HCA)
- Expectation Maximization

○ Visualization and Dimension Reduction

- Principal Component Analysis (PCA)
- Kernel PCA
- Local Linear Embedding (LLE)
- t-distributed Stochastic Neighbor Embedding (t-SNE)

○ Associate Rule Learning

- Apriori
- Eclat

Hierarchy clustering algorithms allow you to subdivide each group into smaller groups. The visualization algorithm creates a 2D or 3D representation that can be schematic by inserting large, unlabeled, high dimensional. Dimension reduction is used to simplify data without losing too much information. For example, the mileage of a car is very associated with the model year, so a dimension reduction algorithm can combine the two characteristics into one characteristic that represents the degree of the car's wear. This is called feature extraction. Abnormal detection is an automatic removal of unusual values from a dataset before injecting them into the learning algorithm.

Reinforcement learning is a very different kind of algorithm. In this case, the learning system is called the agent, and you observe the environment to act and receive rewards. Learn for yourself the best strategy we call policy to get the most rewards over time. Policy is agent is to determine how to behave in a given situation. Deep Mind's Alpha-Go program is also a good example of enhanced learning.

2.2.2. Machine learning algorithm about transportation

The use of machine learning in the areas of transportation varies. The predictions of traffic flow, travel time, real-time traffic density, pedestrian detection and trip routing are as shown in <Table 2-3>.

<Table 2-3> Summary of reviews on machine learning about transportation

Author	Year	Classification of prediction	Machine learning algorithm
Diao et al.	2019	Short term traffic flow	ANN
Dogru et al.	2018	Traffic accident detection	Random Forest
Ma et al.	2015	Real time traffic speed	LSTM-NN
Chung et al	2018	Real time traffic density	Deep-CNN
Kim et al.	2013	Pedestrian detection	Logistic regression

Diao et al. (2019) conducted a study on the prediction of short-term traffic volume using machine learning The study proposed a new hybrid model that accurately predicts the amount of multi-stage forward passenger flow, taking factors into account in terms of time, origin purpose space, frequency and self similarity. For its purpose, first, discrete wavelet transformations were applied to break down the traffic volume series into dedicated and several detailed components. Then, a more efficient tracking model for predicting expenditure elements and a new Gaussian process model for predicting detail were proposed. The predicted performance was evaluated by real-time passenger flow data in Chongqing, China. Simulation results showed that hybrid models could improve accuracy by an average of 20% to 50% especially during rush hours.

Dogru et al. (2018) developed the algorithm of traffic crash detection using RF. This study presented an intelligent traffic crash detection system in which vehicles exchanged minute vehicle variables. The proposed system used simulated data collected from the vehicle's special network (VANET) based on the speed and coordinates of the vehicle, and then transmitted a traffic alert to the driver. It also demonstrated how machine learning algorithms could be utilized to detect crashes occurring on the highways of the ITS. A model was developed to distinguish crash cases from general cases by implementing supervised machine learning algorithms such as ANN, SVM, and RF for traffic data. In terms of accuracy, the performance of the RF algorithm was judged to be superior to that of the ANN and SVM algorithms. RF algorithms performed better with accuracy of 91.56% than ANN with 88.71% SVM with 90.02%.

Ma et al. (2015) developed prediction model of traffic speed using remote microwave sensor data. This study proposed Long Short Term Neural Network (LSTM-NN), a new structure of neural network, to effectively capture nonlinear transport dynamics. LSTM-NN could overcome the problem of back propagated error decay through memory blocks, thus demonstrating excellent ability in time series prediction with long term time dependence. LSTM-NN could also automatically determine the optimum time delay. To verify the effects of LSTM-NN, the moving speed data of the traffic microwave detector in Beijing was used for model training and testing. Comparisons with different topology and other dominant parameters and non-parametric algorithms of dynamic neural networks have shown that LSTM-NN could achieve the best predictive performance in terms of accuracy and stability.

Chung et al. (2018) developed the image based learning methodology to measure traffic density. In this paper, a supervised learning methodology that required no such feature engineering was used. A deep Convolutional Neural Network (CNN) was devised to count the number of vehicles on a road segment based solely on video images. The present methodology did not regard an individual vehicle as an object to be detected separately; rather, it collectively counted the number of vehicles as a human would. The test results showed that the proposed methodology outperformed existing schemes.

Kim et al. (2013) studied a pedestrian detection method using feature selection based on logistic regression analysis. As the parent features, Haar-like and Histograms of Oriented Gradients (HOG) features were used manually. For the statistical analysis, stepwise forward selection, backward elimination, and Least Absolute Shrinkage and Selection Operator (LASSO) methods were applied to Logistic Regression Model for Pedestrian Detection (LRMPD). The results of experiment showed that the average of 48.5% of a full model were selected for LRMPD and this classifier shows performance of up to 95% for detection rate with an approximately 10% false positive rate.

Yu et al. (2010) developed hybrid models based on SVM and Kalman filtering techniques to predict bus arrival times. First of all, using the SVM model, reference travel time was predicted for a given time, weather conditions, path segments, time of movement in the current segment, and the latest time of movement in the predicted segment. In addition, the latest bus arrival information was predicted using the Kalman filtering-based dynamic algorithm. The results showed that the hybrid model proposed in this paper was feasible and applicable in the area of bus arrival time prediction and generally provided better performance than the ANN based method.

2.3. Crash prediction model

2.3.1. Frequency of crashes

Crash prediction models for the frequency of crashes are mostly developed based on GLM such as negative binomial or Poisson function. They have also been developed on the Safety Performance Functions (SPFs) basis. The summary of relevant researches are shown in <Table 2-4>.

<Table 2-4> Summary of reviews on crash frequency

Author	Year	Subject	Methodology
Torre et al.	2019	Development of an accident prediction model	SPF/CMF
Popoola et al.	2017	Accident prediction model on pavement condition and traffic characteristics	(Zero-Inflated) Negative Binomial & Ordered logistic model
Fink et al.	2016	Quantifying the impact of adaptive traffic control systems on crash frequency and severity	Negative binomial & Multinomial logit model
Gianfranco et al.	2018	Accident prediction model for urban road networks	Poisson and Negative binomial regression

Torre et al. (2019) developed an Accident Prediction Model (APM) based on SPFs. APMs represent one of the best tools to perform a road safety quantitative evaluation. This study defined two APMs for single and multiple vehicle fatal-and-injury crashes to be applied on Italian rural freeway segments, based on jurisdictional specific Safety Performance Functions (SPFs) developed in the PRACT project. The proposed procedure was based on the Highway Safety Manual (HSM) approach, and it introduced a new methodology to transfer the HSM to European motorways. In order to improve the prediction accuracy, the proposed APMs consisted in a jurisdictional specific base SPF, developed for the base data set as a function of Annual Average Daily Traffic (AADT) and segment length, combined with Crash Modification Factors (CMFs), in order to account for differences between each site and the base conditions. The full models were then calibrated based on the total number of crashes observed in the wide data set. For both full models (one for single-vehicle and one for multiple-vehicle crashes), the goodness of fit was evaluated in terms of chi square test, root mean square error. The results showed a good aptitude of both models to describe the analysis data set. The proposed models represented a solid and reliable tool for practitioners to perform crash predictions along the Italian freeway network.

Popoola et al. (2017) developed a model for predicting the frequency of crashes on the integration of pavement condition and traffic characteristics in Nigeria. A comparative analysis of the road crash frequency prediction model of the IIesha-Akure-Owo road based on the observed data between 2012 and 2014 was made. Negative Binomial (NB), Ordered Logistic (OL), and Zero Inflated Negative Binomial (ZINB) models were used to model the frequency of crash occurrence using crash data. The explanatory variables included Annual Average Daily Traffic (AADT), Shoulder Factor (SF), Rut Depth (RD), Pavement Condition Index (PCI), and International Roughness Index (IRI). Statistically significant explanatory variables for the three models were AADT, SF, and IRI. The estimated coefficients having the expected signs. Crashes on roads increased with traffic volume and international roughness index, while decreasing with shoulder factors. The systematic variation explained by the models amounts to 87.7%, 78.1%, and 74.4% for NB, ZINB, and OL respectively.

Fink et al. (2016) conducted a study to quantify the impact of adaptive traffic control systems on crashes frequency and severity. This study examined the safety benefits of adaptive traffic control systems using a large SCATS-based system in Oakland County, known as FAST-TRAC. The study used data obtained from FAST-TRAC controlled intersections in Oakland County, comparing similar intersections in other metropolitan areas of Michigan with a wide range of geometric, traffic and collision characteristics. A cross -sectional analysis was performed using data obtained from 498 signalized intersections. The negative binomial model was used to estimate the model for three dependent crashes variables. The multi-nomial logit model was used to estimate the injury severity model. Studies showed that if SCATS-based controllers were at intersections, angular collisions were reduced by up to 19.3%. Severity results showed a statistically significant increase in non- critical injuries, but not a significant decrease in incapacitation or fatal crashes.

Gianfranco et al. (2018) developed an crash prediction model for urban road networks. The study developed a predictive model of urban roads that could estimate the number of crashes for the three situations of urban road networks, detours, three-distance or range bifurcation points, and straight roads. Model development was based on a binary algorithm of Poisson and negative and could be easily applied to crash prediction or the identification of black spots.

2.3.2. Severity of crash

Most of the models that predicted crash severity conducted the research using machine learning algorithms for classification. The summary of relevant researches are shown in <Table 2-5>.

<Table 2-5> Summary of reviews on crash severity

Author	Year	Subject	Methodology
Chang et al.	2006	Analysis of traffic injury severity	CART
Olutayo et al.	2014	Traffic accident analysis	DTs & NN
Alkheder et al.	2016	Severity prediction of traffic accident using an artificial neural network	ANN
Iranitalab et al.	2017	Comparison of four machine learning algorithms for crash severity prediction	MNL, NNC, SVM, RF
Sameen et al	2017	Severity prediction of traffic accidents	RNN
Wang et al.	2017	Analysis of roadway and environment factors affecting traffic crash severities	Logistic regression

Chang et al. (2006) analyzed traffic injury severity using non-parametric classification tree techniques. Statistical regression models, such as logit or probit models, have been widely adopted to analyze the severity of injuries in crashes However, most regression models have their own model assumptions and predefine base relationships between dependent and independent variables. If this assumption is violated, the model may incorrectly estimate the likelihood of injury. On the other hand, The Classification And Regression Tree (CART) does not require a predefined base relationship between the target (dependent) and predictor (independent) variables, and is being used as a powerful tool to deal with predictive and classification issues in particular. In this study, using crash data from 2001 in Taipei, Taiwan, the development of the CART model was carried out to establish the relationship between injury severity and driver/vehicle characteristics, highway/environmental variables and crash variables. The result of study showed that the most important variable related to crash severity was the vehicle type. Pedestrians, motorcycles and cyclists were found to have a higher risk of injury than other types of motorists in crashes.

Olutayo et al. (2014) studied crash analysis using ANN and DT techniques to analyze the causes of crashes on Nigeria's busiest roads. The data were compiled into continuous and categorical data. Continuous data was analyzed using ANN techniques and categorical data was also analyzed using DT techniques. Performance measures used to determine the performance of techniques included instances that were correctly classified as Mean Absolute Error (MAE), confusion matrix, accuracy rate, true positive, false positive and percentage. According to the evaluation results, the DT approach between the machine learning paradigms considered surpassed the ANN with low error rates and high accuracy. It also showed that the three most important causes of the crash were tire rupture, loss of control, and over speeding.

Alkheder et al. (2016) studied severity prediction of crash using an ANN. The model was developed to predict the severity of crashes based on crash records in Abu-Dhabi. An ANN classifier was built using Wikato Environment for Knowledge Analysis (WEKA) data mining software for knowledge analysis. The experimental results showed that the developed ANN classifier could predict the severity of the crash with reasonable accuracy. The overall model's forecast performance was 74.6%. To improve the predictive accuracy of ANN classifiers, crash data were divided into three clusters using k-means algorithms. The post-clustering results showed a significant improvement in the predicted accuracy of the ANN classifier. In this study, the sequential provisioning model was also used as a comparative benchmark to verify the performance of the ANN model. The R tool was used to perform an ordered probit. For each crash, the ordered probit model showed how likely this crash would result in each class (minor, moderate, severe and death). The accuracy of 59.5% obtained from the ordered probit model was clearly less than the ANN accuracy value of 74.6%.

Iranitalab et al. (2017) developed a model that predicted crash severity by applying four statistical and machine learning algorithms. In predicting the severity of crashes, predictive performance was compared using Multi-Nomial Logit (MNL), Nearest Neighbor Classification (NNC), SVM, and RF. In addition, the effects of the method of data clustering consisting of constant prediction, K-means

Clustering (KC), and Latent Class Clustering (LCC) on the performance of the crash severity prediction model were investigated. The four prediction methods were trained/estimated using the training/estimation dataset and the correct prediction rates for each crash severity level, overall correct prediction rate and a proposed crash costs-based accuracy measure were obtained for the validation dataset. Results of study have shown that NNCs had the best predictive performance in overall and more severe collisions. Next, RF and SVM had sufficient performance and MNL was the weakest. Data clustering did not affect the forecast results of the SVM, but KC improved the predictive performance of MNL, NNC, and RF, while LCC resulted in improvements in MNL and RF, but weakened the performance of the NNC.

Samaine et al. (2017) developed an deep learning model that predicted the degree of injury to crashes based on the record of crashes occurring on Malaysia's North-South Expressway using Recurrent Neural Network (RNN). Compared to the traditional Neural Networks (NN), the RNN method was expected to be more effective in sequential data and capture time correlation during crash records. The selected network architecture consisted of a Long Short Term Memory (LSTM) layer, two fully connected (dense) layers, and a Soft-max layer. Next, 0.3 probability dropout technique was applied to avoid over-fitting. In addition, networks were trained in the Tensor -flow framework with Stochastic Gradient Descent (SGD) algorithms (learning rate = 0.01). Additional sensitivity analyses of RNN models were performed to determine the effect of factors on injury severity results. Performance was also evaluated by comparing the proposed RNN model with the Multi Layer Perceptron (MLP) and Bayesian Logistic Regression (BLR). Comparative analysis has shown that the RNN model outperforms the MLP and BLR Validation accuracy of RNN models reached 71.77%, while MLP and BLR models achieved 65.48% and 58.30%, respectively. The results of this study indicated that in a deep learning framework, the RNN model could be a promising tool for predicting the severity of injuries in crashes.

Wang et al. (2017) analyzed road and environmental factors affecting the severity of crashes. This study identified and quantified the effects of several major road and environmental factors on the severity of crashes, and then proposed ways to reduce traffic fatalities and injuries by emphasizing specific road types under certain environmental conditions. A logistic regression model was developed to predict the probability that a crash would cause fatal/serious injury depending on the combination of different roads and environmental conditions. The results of the study showed that the road function class, crash location, road alignment, lighting condition, road surface condition, and speed limit had a significant effect on the severity of traffic collision. The high severity of the impact was associated with rural roads, major arterial roads other than intersection positions, curved positions, dark and dry road conditions without street lights, and high speed limits.

2.4. Interpretable Machine Learning (IML)

2.4.1. Introduction

Machine learning models have demonstrated great success in learning complex patterns that enable them to make predictions about unobserved data. In addition to using models for prediction, the ability to interpret what a model has learned is receiving an increasing amount of attention. However, this increased focus has led to considerable confusion about the notion of interpretability. In particular, it is unclear how the wide array of proposed interpretation methods are related, and what common concepts can be used to evaluate them. In this regard, research on IML techniques that take into account not only the predictive performance of machine learning but also the interpretability has been attempted recently. The studies of IML's definitions, methods, and applications are introduced as shown in <Table 2-6>.

<Table 2-6> Introduction on IML techniques

Author	Year	Subject
Murdoch et al.	2018	Interpretable machine learning; definitions, methods, and applications
Du et al.	2019	Techniques for Interpretable Machine Learning; designing user-friendly explanations and developing comprehensive evaluation metrics
Mohseni et al.	2018	A Survey of Evaluation Methods and Measures for Interpretable Machine Learning

Murdoch et al. (2018) defined interpretability in the context of machine learning and introducing the Predictive, Descriptive, and Relevant (PDR) framework for discussing interpretations. The PDR framework provides three overarching desiderata for evaluation: predictive accuracy, descriptive accuracy, and relevancy, with relevancy judged relative to a human audience. Moreover, to help manage the deluge of interpretation methods, they introduced a categorization of existing techniques into model-based and post-hoc categories, with sub-groups including sparsity, modularity and simulatability.

Du et al. (2019) provided a survey covering existing techniques to increase the interpretability of machine learning models. they also discussed crucial issues that the community should consider in future work such as designing user-friendly explanations and developing comprehensive evaluation metrics to further push forward the area of IML.

Mohseni et al. (2018) proposed the different evaluation goals in interpretable machine learning research by a thorough review of evaluation methodologies used in machine-explanation research across the fields of human-computer interaction, visual analytics, and machine learning. They presented a 2D categorization of IML evaluation methods and showed a mapping between user groups and evaluation measures. Further, they addressed the essential factors and steps for a right evaluation plan by proposing a nested model for design and evaluation of explainable artificial intelligence systems.

2.4.2. Application of IML

IML techniques are actively studied in engineering fields such as energy, logistics, pattern recognition, and medical fields such as diagnosis of disease. Reviews of the IML methodologies applicable to each field are shown in <Table 2-7>.

<Table 2-7> Summary of reviews on IML

Author	Year	Subject	Methodology	
			Black-box model	Interpretable model
Fan et al.	2018	A novel methodology to explain and evaluate data-driven building energy performance models	GLM, MLP, SVM, RF, XGB	LIME
Baryannis et al.	2019	Predicting supply chain risks using machine learning	SVM	Decision Tree
Karatekin et al.	2019	Interpretable Machine Learning in Healthcare : Predicting Severe Retinopathy of Prematurity	DNN	Logistic regression
Xi et al.	2018	Interpretable Machine Learning with labelled handwriting digits	CNN	Fuzzy logic based rule

Fan et al. (2018) proposed a comprehensive methodology to explain and evaluate data-driven building energy performance models. The methodology was developed based on the framework of IML. It can help building professionals to understand the inference mechanism learnt, e.g., why a certain prediction is made and what are the supporting and conflicting evidences towards the prediction. A novel metric was proposed as an alternative approach other than conventional accuracy metrics to evaluate model performance. The methodology has been validated based on actual building operational data. The results obtained were valuable for the development of intelligent and user-friendly building management systems.

Baryannis et al. (2019) proposed a supply chain risk prediction framework using data-driven AI techniques and relying on the synergy between AI and supply chain experts. They then explored the trade-off between prediction performance and interpretability by implementing and applying the framework on the case of predicting delivery delays in a real world multi-tier manufacturing supply chain. Experiment results showed that prioritizing interpretability over performance might require a level of compromise, especially with regard to average precision scores.

Karatekin et al. (2019) investigated the risk factors that lead to severe retinopathy of prematurity using statistical analysis and logistic regression as a form of Generalized Additive Model (GAM) with pair-wise interaction terms (GA2M). In this process, they discussed the trade-off between accuracy and interpretability of these machine learning techniques on clinical data. They also confirmed the intuition of expert neonatologists on a few risk factors, such as gender, that were previously deemed as clinically not significant in RoP prediction.

Xi et al. (2018) developed the IML methodology for recognizing labelled handwriting digits. For this, a CNN learning structure was proposed, with added interpretability-oriented layers, in the form of Fuzzy Logic based rules. This was achieved by creating a classification layer based on a Neural-Fuzzy classifier, and integrating it into the overall learning mechanism within the deep learning structure. Using this structure, one could extract linguistic Fuzzy Logic based rules from the deep learning structure directly, which enhanced the interpretability of the overall system. The classification layer was realized via a Radial Basis Function (RBF) Neural-Network, that was a direct equivalent of a class of Fuzzy Logic-based systems. In this work, the development of the RBF neural-fuzzy system and its integration into the deep-learning CNN was presented. The proposed hybrid CNN RBF-NF structure could form a fundamental building block, towards building more complex deep learning structures with Fuzzy Logic based interpretability. Using simulation results on a benchmark data-driven modelling and classification problem they showed that the proposed learning structure maintained a good level of prediction accuracy (> 96% on unseen data) compared to state-of-the-art CNN deep learning structures, while providing linguistic interpretability to the classification layer.

3. Model Specification

3.1. Analysis of SSES effectiveness

3.1.1. Crashes analysis

Crash data were collected from sections of SSES installed on Korean expressways from 2007 to 2019. When collecting data, the crashes of toll gates, lamps, inter-changes, and rest areas were excluded from the scope of collection because it was difficult to determine due to the effects from installing the SSES.

The analyzed results of the total crashes, EPDO, and casualty crashes before-after installation of SSES using the naive before-after test are shown in <Table 3-1>.

<Table 3-1> Result of crash analysis (naive before-after test)

	Before	After	% Change	t-value
Total crash (annual average)	3.87	2.24	-42.15	2.767^{***}
EPDO (annual average)	26.05	7.65	-70.64	1.674^{*}
Casualty crash (annual average)	1.54	0.84	-45.35	1.833^{*}

$^{*}p<0.1$ $^{**}p<0.05$ $^{***}p<0.01$

Reduction rate of the total crash was 42.15%, that of EPDO was 70.64%, and that of casualty crash was 45.35%. The result of independent sample t-test between the before-after showed that the total crash was statistically significant at the 99% confidence level and the EPDO and the casualty crash were statistically significant at the 90% confidence level.

The analysis results of total crashes, EPDO and casualty crashes before-after installing the SSES using the C-G method are shown in <Table 3-2>.

<Table 3-2> Result of crash analysis (C-G method)

	Total crash (annual average)	EPDO (annual average)	Casualty crash (annual average)
Number of crash in target group (before)	185	26.05	75
Number of crash in target group (after) (λ)	109	7.65	37
Number of crash in comparison group (before)	133	13.65	61
Number of crash in comparison group (after)	115	12.89	34
Number of prediction crash in target group (after) ($\hat{\pi}$)	158.77	22.92	41.13
Reduction in crash (δ)	47.77	15.27	4.13
Effectiveness Index (θ)	0.67	0.28	0.84
Variation rate (%)	-31.35	-66.62	-10.04

※ comparison groups are the same section in the opposite direction of the installation of SSES

Reduction rate of the total crash was 31.35%, that of EPDO was 66.62%, and that of casualty crash was 10.04%. Because all of the effectiveness index (Θ) are smaller than 1, there are the effect of reducing the total crashes, EPDO, and casualty crashes when installing the SSES.

3.1.2. Speed analysis

Speed analysis was carried out through the Vehicle Detection System (VDS) data within a one-year period before-after installation of SSES. The results of the analysis for average speed and proportion of exceeding the speed limit before-after installation of SSES using the naive before-after test are shown in <Table 3-3>.

<Table 3-3> Result of speed analysis (naive before-after test)

	Before	After	% Change	t-value
Average speed	97.57km/h	90.82km/h	-6.92%	4.156^{***}
Proportion of exceeding the speed limit	28.76%	8.26%	-20.50%p	3.388^{***}

**p<0.05, ***p<0.01

Reduction rate of the average speed was 6.92% and proportion of exceeding the speed limit was 20.50p%. The results of independent sample t-test between the before-after showed that both the average speed and proportion of exceeding the speed limit were statistically significant at the 99% confidence level.

In addition, the results of the analysis for the average speed and proportion of exceeding the speed limit before-after installation of SSES using the C-G method are shown in <Table 3-4>. Reduction rate of the average speed was 3.49% and proportion of exceeding the speed limit was 56.65%. Because all of the effectiveness index (Θ) are smaller than 1, there are the effects of reducing the average speed and proportion of exceeding the speed limit when installing the SSES.

<Table 3-4> Result of speed analysis (C-G method)

	Average speed	Proportion of exceeding the speed limit
Speed in target group (before)	97.57km/h	28.76%
Speed in target group (after) (λ)	90.82km/h	8.26%
Speed in comparison group (before)	98.96km/h	35.03%
Speed in comparison group (after)	96.41km/h	23.87%
Prediction Speed in target group (after) ($\hat{\pi}$)	94.10	19.05
Reduction in speed (δ)	3.28	10.79
Effectiveness Index (θ)	0.93	0.39
Variation rate (%)	-3.49	-56.65

※ comparison groups are 2km of the upper and lower sections of the installation section of SSES.

The speed analysis showed that the installation of SSES greatly affected the reduction proportion of exceeding the speed limit rather than the average speed, since it reduced the speed of the vehicle below the speed limit. It was also found that the rate of speed reduction using the C-G method was less effective than that of the naive before-after test.

3.2. Data collection & pre-analysis

3.2.1. Data collection

As of 2019, SSESs are installed and being operated in 97 nationwide sections. Data were collected on Korean expressways, where data needed for the development of the crash prediction model was collected. The scope of data collection is shown in <Table 3-5>. The temporal scope of data collection is from 2010 to 2019, the spatial scope is the sections of SSES installation and same sections in the opposite direction of the SSES installation. The content scope is crash data, road conditions, traffic conditions, and control conditions.

<Table 3-5> Scope of data collection

Scope	Data collection
Temporal	· 2010 ~ 2019
Spatial	· Sections of SSES installed in the Korean expressway · Same sections in the opposite direction of the SSES installed
Content	· Crashes, speed, traffic volume, road conditions, etc.

The contents of data collection are shown in <Table 3-6>. The number of lanes, entry and exit, the length or ratio of slopes, tunnels, bridges were collected as the road conditions. Traffic volume, heavy vehicle ratio, average speed, speed variation, and the proportion of exceeding the speed limit were collected as the traffic conditions. And speed limit, whether or not SSES installation were as the control conditions.

<Table 3-6> Contents of data collection

Contents	Variables
Road condition	No. of lanes, slopes, curves, tunnels(number/length), bridges(number/length), No. of entry/exit
Traffic condition	Traffic volume, heavy vehicle ratio, average speed, speed variation, proportion of exceeding the speed limit
Control condition	Speed limit, SSES(O/X), length of SSES section

3.2.2. Basic statistics of variables

First of all, a basic analysis of the relationship between SSES and casualty crash was conducted prior to quantifying the effect of reducing the casualty crash due to the installation of SSES, which was the objective of this study. As shown in the <Table 3-7>, cross table was drawn on whether or not SSES installation and whether or not a casualty crash occurrence.

<Table 3-7> Contingency table between SSES and casualty

		SSES		Total
		0	1	
Casualty	0	88	70	158 (56.4%)
	1	91	31	122 (43.6%)
Total		179 (63.9%)	101 (36.1%)	280 (100%)

Of the total data, 36.1% of the sections was installed with SSES and 63.9% of the sections was not. In addition, 43.6% of the sections where the casualty crash occurred, and 56.4% of the sections that were not occurred.

Secondly, basic statistics were analyzed for the collected variables, such as speed, traffic volume, and geometry. The definition and description of collected variables are as shown in <Table 3-8>.

<Table 3-8> Variable description

Variable	Definition	Description
MS	Mean Speed	Average of the all's VDS speed every 5 minutes within the section
SV	Speed Variance	Speed variance between average speeds of each VDS every 5 minutes within the section
SOR	Speed Over Ratio	The proportion of time exceeding the speed limit among all's VDS speed every 5 minute within the section
TVL	Traffic Volume Lane	ln(annual average daily traffic volume per lane) within the section
HVR	Heavy Vehicle Ratio	The ratio of trucks more than 2.5t or buses more than 16 passengers
BR	Bridge Ratio	(Total length of bridge(s) within section/section length) *100
TR	Tunnel Ratio	(Total length of tunnel(s) within section/section length) *100
CR	Curve Ratio	(Total length of curve(s) within section/section length) *100 ※ curve: side slope percentage is more than 3%
SR	Slope Ratio	(Total length of slope(s) within section/section length)*100 ※ slope: upward or downward slope percentage is more than 2%
SL	Speed Limit	Speed limit within the section
LS	Length Section	The length from start point to end point of SSES
L	Lane	Number of lanes

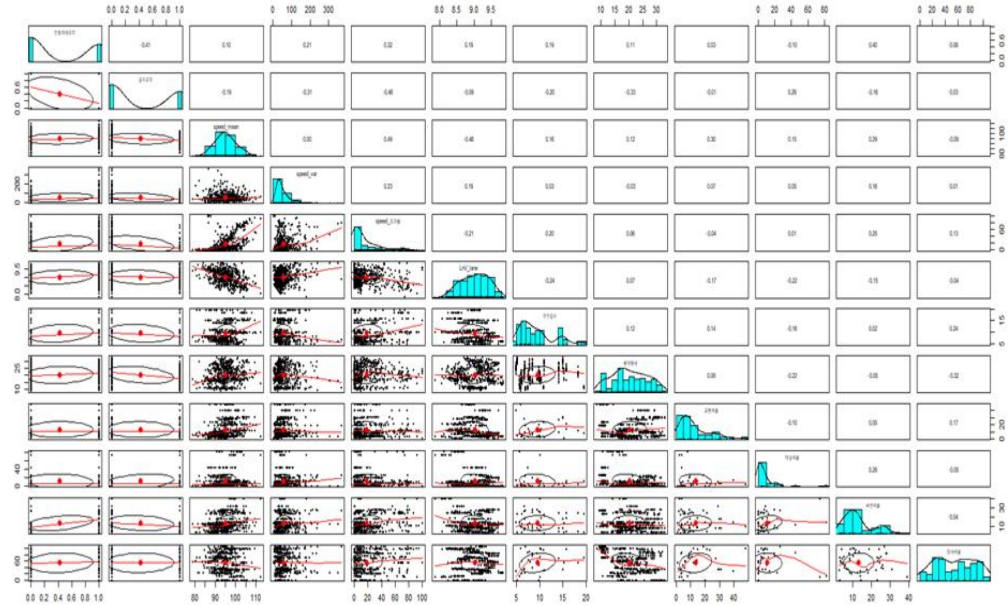
※ Section: area from start point to end point of SSES

The results of analysis for minimum, maximum, average, and standard deviation of variables are shown in <Table 3-9>.

<Table 3-9> Basic statistics value

Variable	Min	Max	Mean	Stdev
MS	81.45	112.24	95.82	6.24
SV	4.19	368.05	67.74	59.19
SOR	0.00	99.73	27.08	30.91
TVL	7.49	10.90	9.14	0.65
HVR	9.35	35.23	20.04	5.78
BR	0.86	47.92	12.74	10.61
TR	0.00	82.14	11.76	19.41
CR	2.58	39.11	11.59	8.81
SR	0.00	47.19	31.70	21.26
SL	80	110	101.79	5.65
LS	4.90	19.50	9.98	3.81
L	2	4	2.38	0.74

Thirdly, A scatter plot between variables is drawn in [Figure 3-1]. It is a type of plot or mathematical diagram using cartesian coordinates to display values for typically two variables for a set of data. The data are displayed as a collection of points, each having the value of one variable determining the position on the horizontal axis and the value of the other variable determining the position on the vertical axis.



[Figure 3-1] Scatter plot between variables

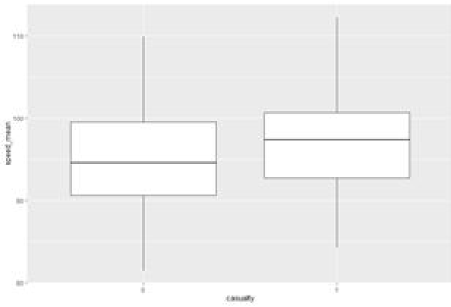
Also, a correlation analysis was conducted as shown in <Table 3-10>. It is a numerical measure of some type of correlation, meaning a statistical relationship between two variables. The variables may be two columns of a given data set of observations, often called a sample, or two components of a multi-variate random variable with a known distribution. Several types of correlation coefficient exist, each with their own definition and own range of usability and characteristics. They all assume values in the range from 1 to +1, where ±1 indicates the strongest possible agreement and 0 the strongest possible disagreement.

<Table 3-10> Correlation coefficient between variables

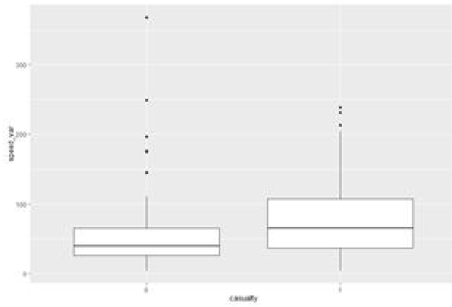
	casualty	SSES	SM	SV	SOR	TVL	HVR	BR	TR	CR	SR	SL	LS	L
casualty	1	-.411**	.155**	.177**	.280**	.185**	0.096	0.070	-0.028	0.064	0.042	-0.023	0.061	-0.002
SSES		1	-.321**	-.229**	-.416**	0.027	-0.098	-0.003	-0.003	0.002	-0.012	0.016	-0.025	0.025
SM			1	-0.061	.566**	-.326**	-0.034	.173**	.399**	.229**	-.152*	.234**	.120*	-.210**
SV				1	-0.023	.174**	-.201**	0.084	.217**	.196**	0.065	.215**	-0.059	0.046
SOR					1	-.124*	-.169**	-.183**	.378**	.137*	.129*	-.504**	.238**	-.178**
TVL						1	-0.012	-0.045	-.142*	0.012	0.108	-0.044	-.316**	.209**
HVR							1	0.064	-.262**	-.299**	-.228**	0.114	.148*	-.135*
BR								1	-.159**	0.108	.121*	.463**	0.086	-.405**
TR									1	.507**	-.148*	-0.092	-.157**	-.252**
CR										1	-0.067	0.063	-.322**	-.145*
SR											1	-.348**	.130*	-.191**
SL												1	-0.112	0.043
LS													1	-.262**
L														1

*p<0.1 **p<0.05 ***p<0.01

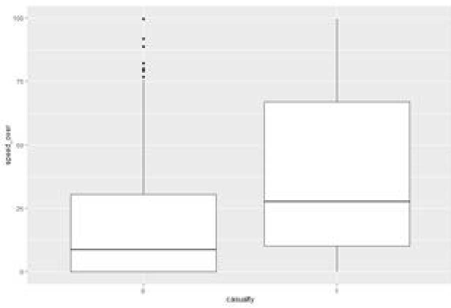
A box plot is a method for graphically depicting groups of numerical data through their quartiles. Box plots may also have lines extending from the boxes (whiskers) indicating variability outside the upper and lower quartiles, hence the terms box-and-whisker plot and box-and-whisker diagram. Outliers may be plotted as individual points. Box plots are non-parametric and they display variation in samples of a statistical population without making any assumptions of the underlying statistical distribution. The spacings between the different parts of the box indicate the degree of dispersion and skewness in the data, and show outliers. In addition to the points themselves, they allow one to visually estimate various L-estimators, notably the inter-quartile range, mid-hinge, range, mid-range, and tri-mean. The results of the bot-plot between the major independent variables and the occurrence of a casualty crash are shown in the following [Figure 3-2].



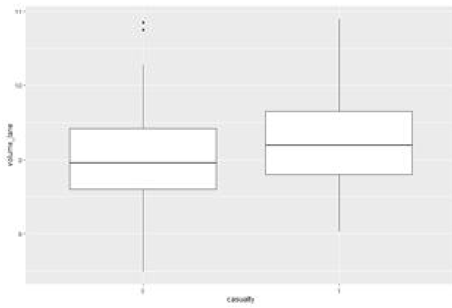
<SM>



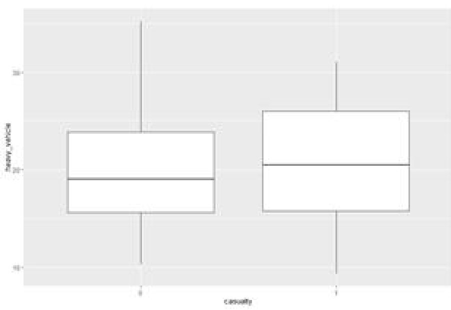
<SV>

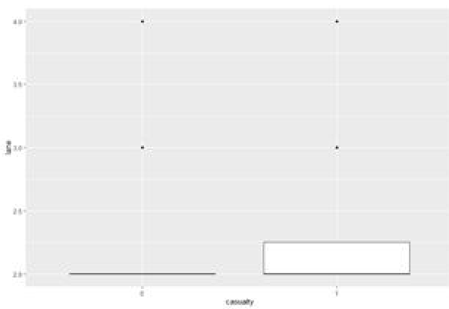


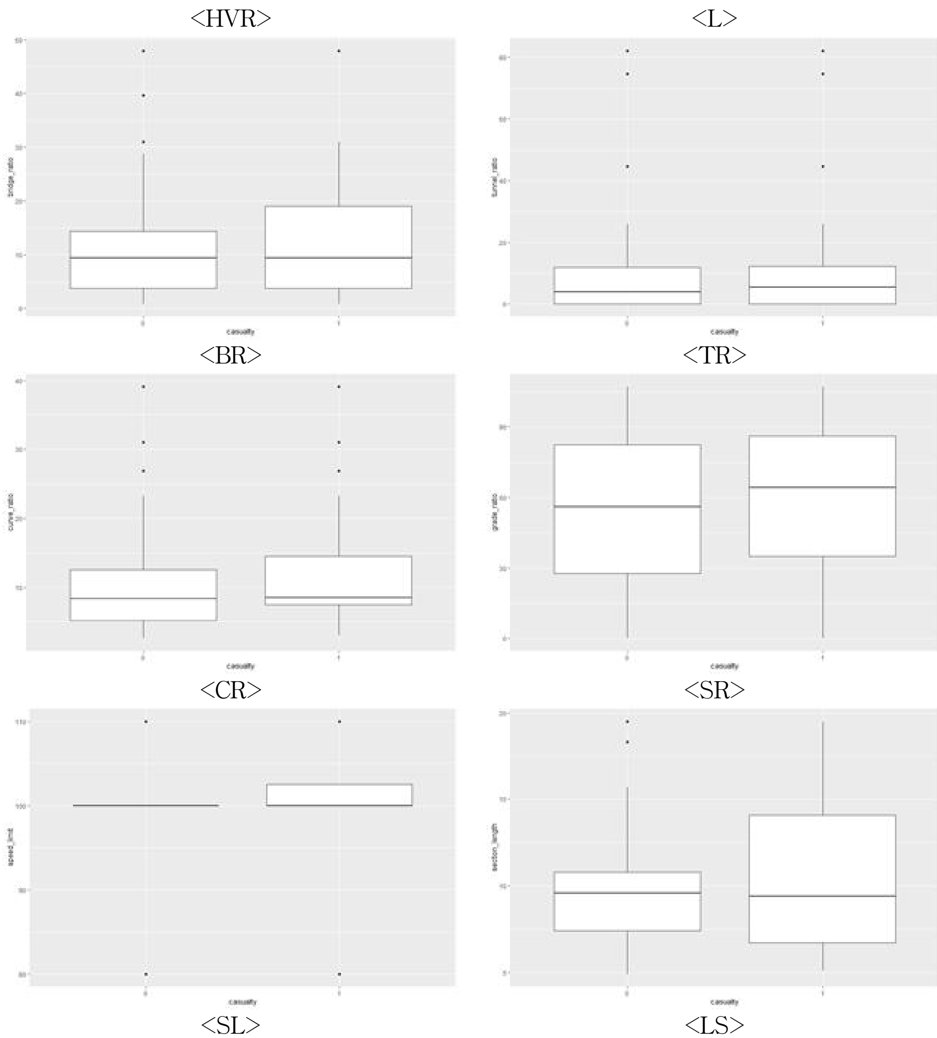
<SOR>



<TVL>







[Figure 3-2] Box and whisker plot

Also, the independent sample t-test between the major independent variables and the occurrence of a casualty crash were conducted. The SM, SV, SOR, and TVL of the independent variables were found to be statistically significant in the least 95% confidence level and the other variables were not statistically significant. Casualty crash can be judged to be significantly affected by related variables with speed and traffic volume. Independent sample t-test results are shown in the following <Table 3-11>.

<Table 3-11> t - test results for casualty

Casualty	SM	SV	SOR	TVL	HVR	BR	TR	CR	SR	SL	LS	L
0	94.97	58.57	19.48	9.04	19.55	12.10	12.23	11.10	55.53	101.90	9.78	2.38
1	96.92	79.61	36.92	9.28	20.66	13.58	11.14	12.23	58.23	101.64	10.25	2.38
t-value	-2.611**	-2.991***	-4.867***	-3.137***	-1.605	-1.162	0.467	-1.068	-0.708	0.380	-1.023	0.030

**p<0.05, ***p<0.01

The similar process was conducted considering that whether or not SSES was installed. SM, SV, and SOR of the independent variables were found to be statistically significant in the 99% confidence level and the other variables were not statistically significant. It could be seen that variables related to speed are greatly reduced by the installation of SSES. The results of the independent sample t-test between the independent variables and whether or not the SSES installation are shown in <Table 3-12>.

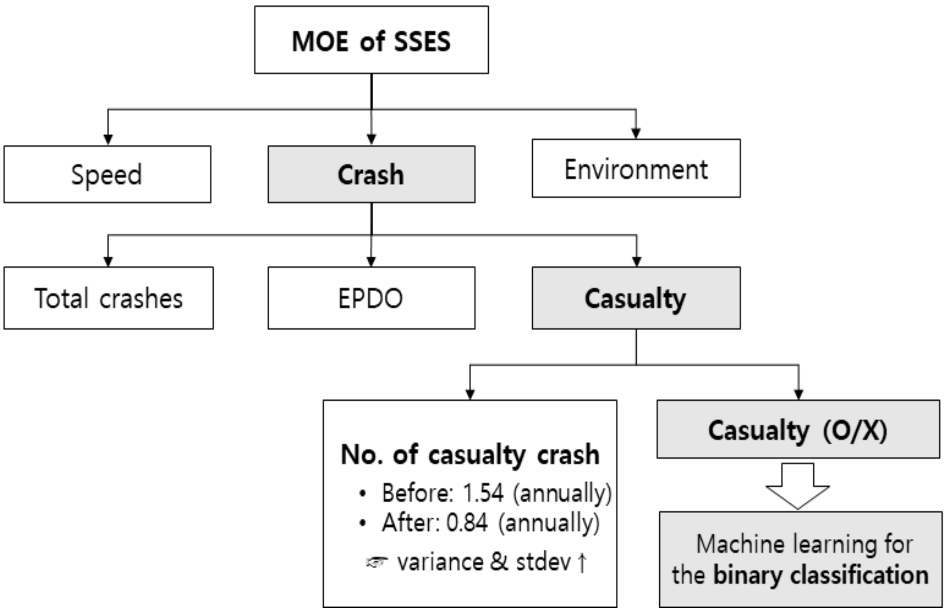
<Table 3-12> t - test results for SSES

SSES	SM	SV	SOR	TVL	HVR	BR	TR	CR	SR	SL	LS	L
0	97.24	77.37	36.21	9.13	20.44	12.77	11.80	11.58	56.97	101.72	10.05	2.37
1	93.01	48.68	9.03	9.17	19.24	12.70	11.67	11.62	56.19	101.91	9.85	2.40
t-value	5.654***	3.929***	7.627***	-0.444	1.649	0.053	0.051	-0.031	0.194	-0.272	0.411	-0.411

**p<0.05, ***p<0.01

3.3. Response variable selection

As mentioned in the chapter 1, the KNPA is installing the SSES for the purpose of reducing casualty crashes through speed control. In this regard, this study conducted the response variable selection process for model development with MOEs used in the analysis of SSES installation effects as considered in the chapter 2 literature review. The selection process of response variable is as shown in [Figure 3-3].



[Figure 3-3] Response variable selection process

According to a study by Cassetta et al. (2011), variables related to crash, speed, and environment are used as MOEs for analysis of the effectiveness of SSES. The main purpose of installing SSES is to reduce crashes by control the speed. Variables (e.g. mean speed, speed variation, ratio of exceeding the speed limit) related to speed act as the mediation effects in reducing crashes through speed control. Also variables related to environment such as CO_2 emissions may be subordinate effects of installing the SSES.

Therefore, the crash is selected as the primary response variable. The variables related to crash can be divided into the number of crashes, EPDO, and casualties. Total crashes are difficult to represent the purpose of SSES installation which is to reduce casualties through speed control because they contain PDO crashes which are not related to speeding. In the case of EPDO, it is likely to have significant distortion in its prediction of the effects if a major crash including buses and trucks occurs. Therefore, casualty crashes were selected as a response variable in this study. According to the results of effect analysis for SSES, the number of casualty crashes before installing SSES was 1.54 (annually) and after installing SSES was 0.84 (annually). And their variation and standard deviation were large when they were compared with average of casualty crashes. Therefore, it was judged to be possible to apply the binary classification technique to predict the occurrence of casualty crash, not the number of casualty crashes.

Through such a selection process, the occurrence of casualty crash was selected as the response variable in this study and machine learning algorithms of binary classification were used to develop the model.

3.4. Model selection

3.4.1. Binary classification

Binary classification is the task of classifying the elements of a given set into two groups on the basis of a classification rule. Contexts requiring a decision as to whether or not an item has some qualitative property, some specified characteristic, or some typical binary classification include. Binary classification is dichotomization applied to practical purposes, and in many practical binary classification problems, the two groups are not symmetric – rather than overall accuracy, the relative proportion of different types of errors is of interest. For example, in medical testing, a false positive (detecting a disease when it is not present) is considered differently from a false negative (not detecting a disease when it is present).

Statistical classification is a problem studied in machine learning. It is a type of supervised learning, a method of machine learning where the categories are predefined, is used to categorize new probabilistic observations into said categories. When there are only two categories, the problem is known as statistical binary classification.

Some of the methods commonly used for binary classification are:

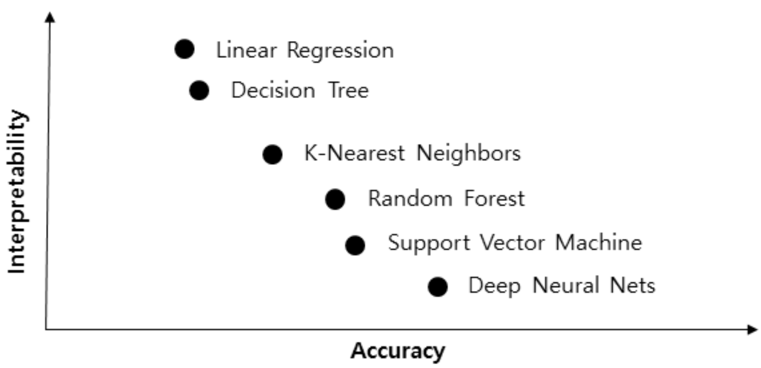
- Decision Trees (DT)
- Random Forest (RF)
- K-Nearest Neighbors (KNN)
- Bayesian networks
- Support Vector Machines (SVM)

- Deep Neural Networks (DNN)
- Binary Logistic Regression (BLR)
- Probit model

Each classifier is best in only a select domain based upon the number of observations, the dimensionality of the feature vector, the noise in the data and many other factors. For example, RFs perform better than SVM classifiers for 3D point clouds.

3.4.2. Accuracy vs. Interpretability

The relation between the accuracy and interpretability capabilities of machine learning models is the friction between being able to accomplish complex knowledge tasks and understanding how those tasks are accomplished. Knowledge vs. Control, Performance vs. Accountability, Efficiency vs. Simplicity, and so on pick your favorite dilemma and they all can be explained by balancing the tradeoffs between accuracy and interpretability.



[Figure 3-4] Relation between interpretability and accuracy

Many machine learning algorithms are complex in nature and, although they result very accurate in many scenarios, they can become difficult to interpret. The correlation between accuracy and interpretability of the well known machine learning algorithms can be shown in [Figure 3-4].

3.4.3. Overview of IML

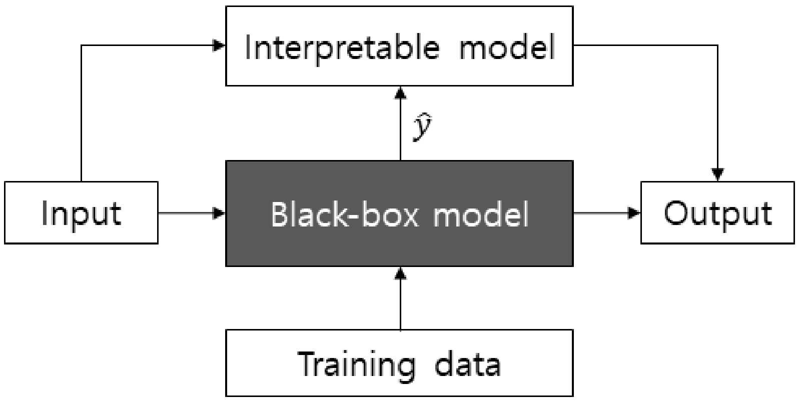
Machine learning is proceeding at an alarming rate by complex models such as ensemble models and DNN. These models range from real life applications such as Netflix's movie recommendations, Google's translation and Amazon's Alexa's voice recognition. In spite of its success, machine learning has its limitations and disadvantages. Most important is the lack of transparency behind their actions, which leaves users with little understanding of how specific decisions are made by these models. For example, a self-driving car with various machine learning algorithms does not brake or decelerate when confronted with a stationary fire engine. This unexpected behavior can frustrate and confuse users, so they can wonder why. Worries about the black-box characteristics of complex models have hindered their further application in our society, especially in important decision making areas such as self-driving cars (Du et al. 2020).

IML is an effective tool to reduce these problems. It gives a machine learning model the ability to explain their behavior in terms that are understandable to humans, which is called interpretability or explainability (Doshi-Velez et al. 2017). Interpretability will be an integral part of the machine learning model to better serve humans and bring benefits to society. For end users, the explanation will encourage increased reliability of a machine learning system. From the perspective of machine learning system developers, the explanations provided can help them better understand why models fail.

IML techniques can generally be divided into two categories: intrinsic interpretability and post-hoc interpretability, depending on the time they are acquired. Intrinsic interpretability is achieved by constructing self-explanatory models which incorporate interpretability directly to their structures. The models of this category include DT, rule-based model, linear model, and attention model. In contrast, post-hoc requires the creation of a second model that provides a description of the existing model. The main difference between these two groups lies in the trade-off between model accuracy and explanation fidelity. Essentially interpretable models can provide accurate and inconsistent explanations, but they can cost some predictive performance. Post-hoc has limitations on approximate nature while retaining the accuracy of the underlying model (Molnar 2018).

IML is further distinguished by two types of interpretability: global interpretability and local interpretability. Global interpretability means that users can understand how the model works globally by examining the structure of a complex model, and local interpretability examines the individual predictions of the model locally to determine why the model makes the decision. These two types bring different benefits. Global interpretability can enhance transparency by shedding light on the internal mechanism of machine learning models. Local interpretability can help to identify the causal relationship between a particular input and its model predictions.

IML consists of a black-box model and an interpretable model, as shown in the [Figure 3-5]. The structure of IML is that prediction results of black-box model with high accuracy performance are interpreted by interpretable model with high explainable performance.



[Figure 3-5] Interpretable machine learning

Black-box models such as DNN, RF, or SVM often provide great accuracy. The inner workings of these models are harder to understand and they don't provide an estimate of the importance of each feature on the model predictions, nor is it easy to understand how the different features interact. Whereas interpretable models such as BLR or DTs on the other hand provide less predictive capacity and are not always capable of modelling the inherent complexity of the dataset (i.e. feature interactions). They are however significantly easier to explain and interpret.

In Carvalho (2019)'s study, these interpretable models are grouped according to the purpose of explanation, as shown in <Table 3-13>.

<Table 3-13> Summary of interpretable models classification (Carvalho et al. 2019)

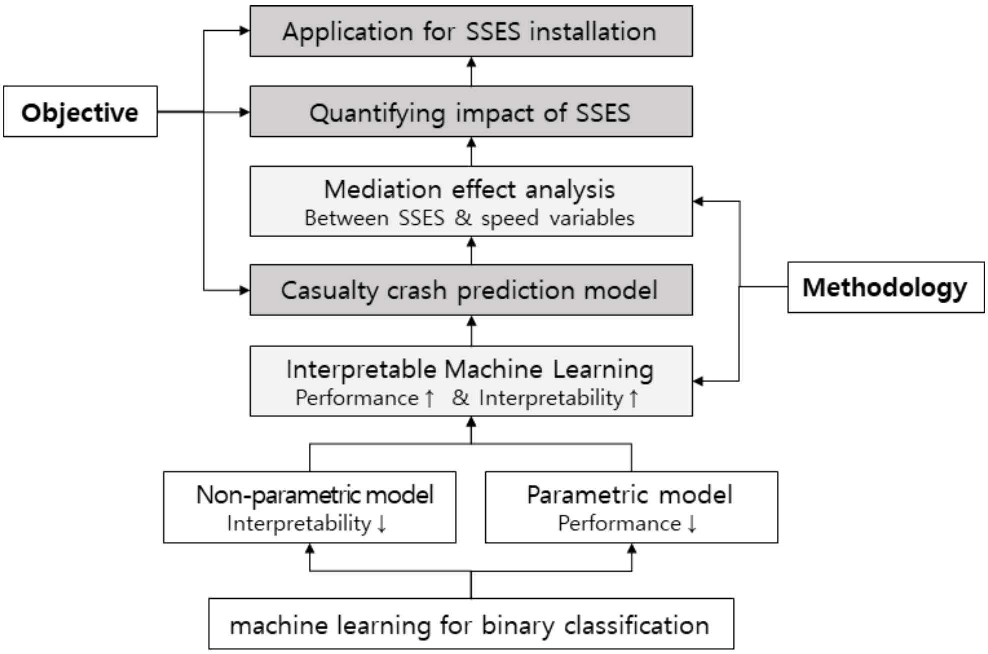
Classification	Content
Interpretability importance	Satisfy human curiosity; Scientific findings; Find meaning Regulation requirements; Social acceptance and trust; Safety Acquire new knowledge
Taxonomy of interpretability	Pre-model vs. In-model vs. Post-model Intrinsic vs. Post-hoc Model-specific vs. Model-agnostic
Scope of interpretability	Algorithm transparency Global model interpretability (holistic vs. modular) Local model interpretability (single vs. group of predictions)
Properties of explanation methods	Expressive power; Translucency; Portability; Algorithmic complexity
Properties of explanations	Accuracy; Fidelity; Consistency; Stability; Comprehensibility; Certainty; Importance; Novelty; Representativeness
Human-friendly explanations	Contrastiveness; Selectivity; Social; Focus on the abnormal; Truthful; Consistent with prior beliefs; General and probable
Interpretability evaluation	Application-level; Human-level; Functional-level
Interpretability goals	Accuracy; Understandability; Efficiency

3.4.4. Process of model specification

In this study, based on the analysis results of SSES installation effects, the occurrence of casualty crash was finally selected as the response variable and machine learning methodology for binary classification was used to develop the casualty crash prediction model.

Non-parametric models with higher accuracy but lower interpretability and parametric models with higher interpretability but lower accuracy were considered for binary classification machine learning algorithms. To overcome the shortcomings of these two categorial models, the IML methodology was applied to develop a predictive model for casualty crash. In addition, the effect of SSES installation was quantified by performing a mediation effect analysis between SSES and variables related to speed. Finally, the criteria for installation of SSES were proposed using a casualty crash prediction model using the IML methodology.

A summary of the model specification process is shown in the following [Figure 3-6].



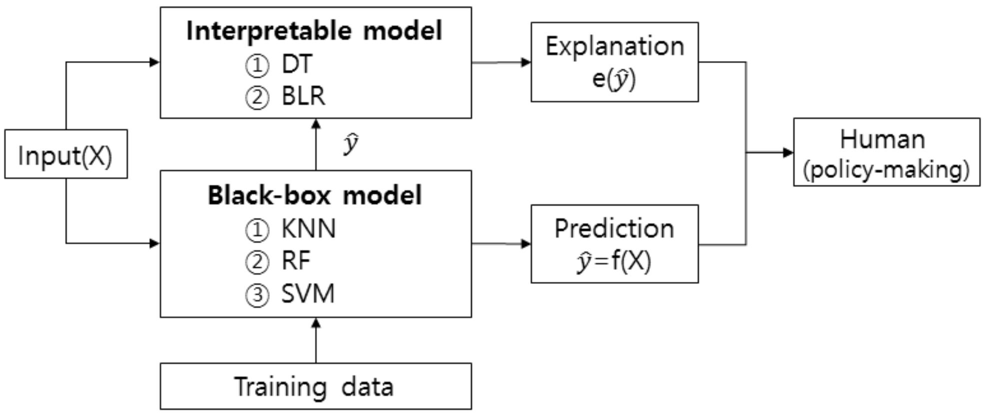
[Figure 3-6] Process of model specification

4. Model development

4.1. Black-box and interpretable model

4.1.1. Consists of IML

IML consists of a black-box model and an interpretable model as shown in [Figure 4-1]. The structure of IML is that prediction result of black-box model with high accuracy performance are interpreted by interpretable model with high explainable performance. In this study, KNN, RF, and SVM were considered as black-box model to increase the accuracy performance, and DT and BLR were considered as interpretable model to increase the interpretability for IML model which predicted the occurrence of casualty crash.



[Figure 4-1] Model selection for IML

The interpretable model is a model that describes the predictive results of a highly predictable black-box model in terms of a human perspective. The process of obtaining interpretable model is as follows (Molnar, 2018).

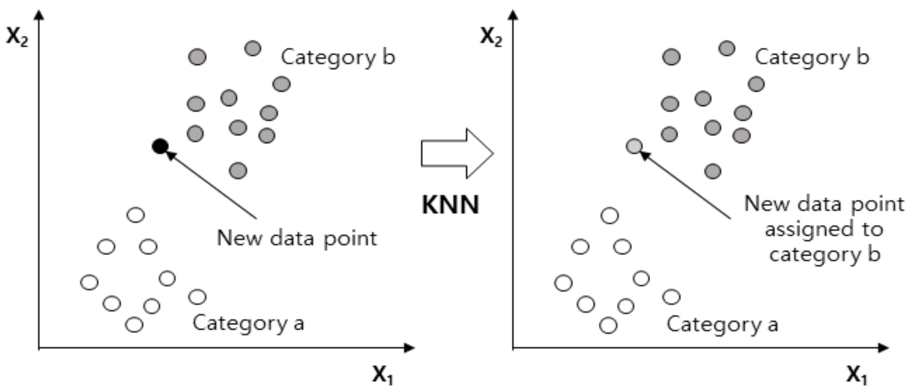
- ○ step-1 Choose a dataset X. This could be the same dataset that was used for training the black-box model or a new dataset from the same distribution. You could even choose a subset of the data or a grid of points, depending on your application.
- ○ step-2 For the chosen dataset X, get the predictions $\hat{y}$ of the black-box model.
- ○ step-3 Choose an interpretable (surrogate) model.
- ○ step-4 Train the interpretable model on the dataset X and its predictions $\hat{y}$.
- ○ step-5 You now have a surrogate model.
- ○ step-6 Measure how well the surrogate model replicates the prediction of the black-box model.
- ○ step-7 Interpret / visualize the surrogate model.

4.1.2. Black-box model

1) KNN

KNN is a non-parametric method used for classification and regression. KNN makes no assumptions about the functional form of the problem being solved (Altman, 1992). In both cases, the input consists of the k closest training examples in the feature space. The output depends on whether KNN is used for classification or regression. In KNN classification, the output is a class membership. An object is classified by a plurality vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small). Whereas in KNN regression, the output is the property value for the object. Its value is the average of the values of k nearest neighbors.

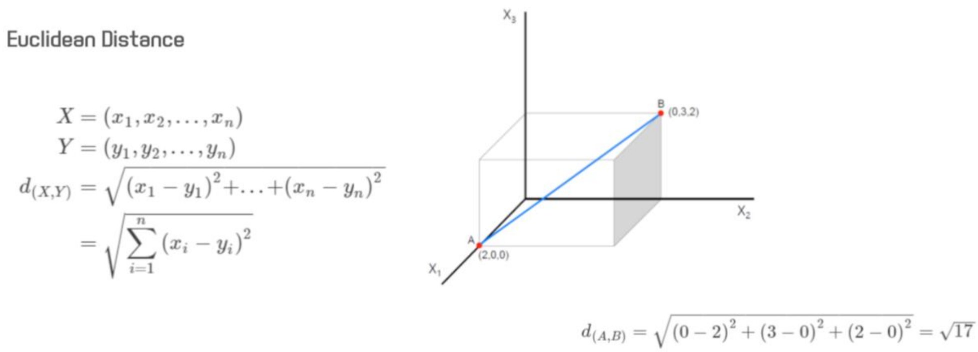
KNN is a type of instance-based learning, or lazy learning, where the function is only approximated locally and all computation is deferred until function evaluation. KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much similar to the new data. The concept of KNN classification is shown in [Figure 4-2].



[Figure 4-2] Concept of KNN classification

There are two main hyper parameters that KNN has to set up to find the best performance. The first is the distance to represent the distance between data and the second is the value of K to be specified by the algorithm.

The distance between data in the KNN model is an important indicator and variable. Because of depending on how you measure distances and set criteria, the classification of new data is different. Commonly used as a way to get the distance are Euclidean's, Manhattan's, Hamming's, and so on. In this study, the Euclidean distance, which is the most commonly used distance calculation, is used. It can get through the distance between two points in the n-dimensional, as shown in [Figure 4-3].



[Figure 4-3] Euclidean distance

The next important hyper-parameter is the K value. K value means how many neighbors to participate in the KNN algorithm. From a model's conformance perspective, it can determine whether the model is over-fitting or under-fitting. If the k value is too small, the classification criteria will be too much strict, so that the accuracy in the train data is high, but the results of high error and low accuracy in the test data can be displayed. In other words, it can be an over-fitting model. On the other hand, if the k value is too large, the classification criteria may be too much general, which makes it less accurate to test data because it is not accurate to the classification of new data. It can be an under-fitting model.

The KNN algorithm is performed according to the following process:

○ Step-1 Select the number K of the neighbors

○ Step-2 Calculate the Euclidean distance of K number of neighbors

○ Step-3 Take the K nearest neighbors as per the calculated Euclidean distance.

○ Step-4 Among these k neighbors, count the number of the data points in each category.

○ Step-5 Assign the new data points to that category for which the number of the neighbor is maximum.

○ Step-6 KNN model is ready.

Advantages and disadvantages of KNN are following:

○ Advantages

- KNN does not learn anything in the training period.
- New data can be added seamlessly.
- KNN is very easy to implement.

○ Disadvantages

- KNN does not work well with large dataset.
- KNN does not work well with high dimensions.
- It is needed to do feature scaling (normalization) before applying KNN algorithm to any dataset.
- KNN is sensitive to noisy data, missing values and outliers.

2) RF

RF is machine learning algorithm that fits many CART models to random subsets of the input data and uses the combined result for prediction (Breiman, 2001). RF is a supervised learning algorithm which is used for both classification as well as regression. But however, it is mainly used for classification problems. RF algorithm creates DTs on data samples and then gets the prediction from each of them and finally selects the best solution by means of voting. It is an ensemble method which is better than a single DT because it reduces the over-fitting by averaging the result.

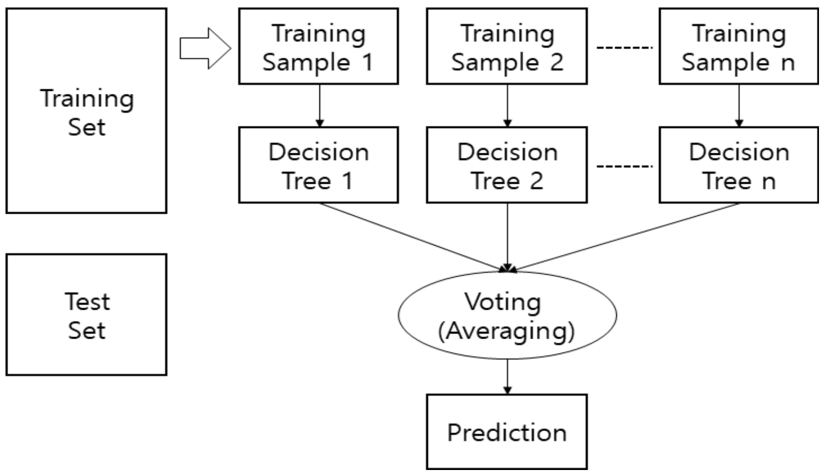
The RF algorithm works as the following [Figure 4-4] and is performed according to the following process:

○ Step-1 Start with the selection of random samples from a given dataset.

○ Step-2 This algorithm will construct a DT for every sample. Then it will get the prediction result from every DT.

○ Step-3 Voting will be performed for every predicted result.

○ Step-4 Select the most voted prediction result as the final prediction result.



[Figure 4-4] Working of RF algorithm

The final result of model is calculated by averaging over all predictions from these sampled trees or by majority vote.

Advantages and disadvantages of RF are following:

○ Advantages

- RF overcomes the problem of over-fitting by averaging or combining the results of different DTs.
- RF works well for a large range of data items than single DT.
- RF has less variance then single DT.
- RFs are very flexible and possess very high accuracy.
- Scaling of data does not require in RF algorithm. It maintains

good accuracy even after providing data without scaling.

- RF algorithms maintains good accuracy even a large proportion of the data is missing.

○ Disadvantages

- Complexity is the main disadvantage of RF algorithms.
- Construction of RF is much harder and more time-consuming than DT.
- More computational resources are required to implement RF algorithm.
- It is less intuitive in case when we have a large collection of DT.
- The prediction process using RFs is very time-consuming in comparison with other algorithms

3) SVM

SVM is supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on the side of the gap on which they fall.

In addition to performing linear classification, SVM can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

Advantages and disadvantages of SVM are following:

○ Advantages

- SVM is very good when there is no idea on the data.
- SVM works well with even unstructured and semi structured data like text, images and trees.
- The kernel trick is real strength of SVM. With an appropriate kernel function, it can solve any complex problem.
- Unlike in neural networks, SVM is not solved for local optima.
- It scales relatively well to high dimensional data.
- SVM models have generalization in practice, the risk of over-fitting is less in SVM.
- When compared to ANN models, SVM gives better results.

○ Disadvantages

- Choosing a good kernel function is not easy.
- Long training time for large data-sets is needed.
- It is difficult to understand and interpret the final model, variable weights and individual impact.
- Since the final model is not so easy to see, we can not do small calibrations to the model hence its tough to incorporate

our business logic.

- The SVM hyper-parameters are cost and gamma. It is not that easy to fine-tune these hyper-parameters. It is hard to visualize their impact.

4.1.3. Interpretable model

1) DT

DT is a decision support tool that uses a tree-like model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm that only contains conditional control statements. DTs are commonly used in operations research, specifically in decision analysis, to help identify a strategy most likely to reach a goal, but are also a popular tool in machine learning.

DT is a flowchart-like structure in which each internal node represents a "test" on an attribute (e.g. whether a coin flip comes up heads or tails), each branch represents the outcome of the test, and each leaf node represents a class label. The paths from root to leaf represent classification rules.

In decision analysis, DT and the closely related influence diagram are used as a visual and analytical decision support tool, where the expected values of competing alternatives are calculated.

The DT can be linearized into decision rules, where the outcome is the contents of the leaf node, and the conditions along the path form a conjunction in the if-clause. In general, the rules have the form:

- if condition 1 and condition 2 and condition 3 then outcome.

Decision rules can be generated by constructing association rules with the target variable on the right. They can also denote temporal or causal relations.

Advantages and disadvantages of DT are following:

○ Advantages

- DT is simple to understand, interpret and visualize.
- DT implicitly performs variable screening or feature selection.
- DT can handle both numerical and categorical data. Can also handle multi-output problems.
- DT requires relatively little effort for data preparation.
- Non-linear relationships between parameters do not affect tree performance.

○ Disadvantages

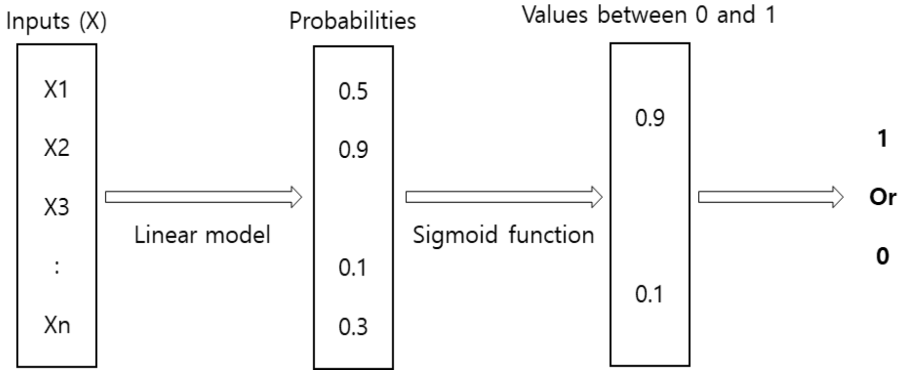
- DT learners can create over-complex trees that do not generalize the data well. This is called over-fitting.
- DT can be unstable because small variations in the data might result in a completely different tree being generated. This is called variance, which needs to be lowered by methods like bagging and boosting.
- Greedy algorithms can't guarantee to return the globally optimal DT. This can be mitigated by training multiple trees,

where the features and samples are randomly sampled with replacement.

- DT learners create biased trees if some classes dominate. It is therefore recommended to balance the data set prior to fitting with the DT.

2) BLR

In statistics, the logistic model (or logit model) is used to model the probability of a certain class or event existing such as pass/fail, win/lose, alive/dead or healthy/sick. This can be extended to model several classes of events such as determining whether an image contains a cat, dog, lion, etc. Each object being detected in the image would be assigned a probability between 0 and 1 and the sum adding to one. BLR measures the relationship between the response variable and the one or more independent variables, by estimating probabilities using it's underlying logistic function. These probabilities must then be transformed into binary values in order to actually make a prediction. This is the task of the logistic function, also called the sigmoid function. The sigmoid function is an S-shaped curve that can take any real-valued number and map it into a value between the range of 0 and 1, but never exactly at those limits. This values between 0 and 1 will then be transformed into either 0 or 1 using a threshold classifier. [Figure 4-5] illustrates the steps that logistic regression goes through to get desired output.



[Figure 4-5] Working of BLR algorithm

Advantages and disadvantages of BLR are following:

○ Advantages of BLR

- BLR is a widely used technique because it is very efficient, does not require too many computational resources,
- BLR is highly interpretable, it does not require input features to be scaled, it does not require any tuning, it's easy to regularize, and it outputs well-calibrated predicted probabilities.
- BLR does work better when removing attributes that are unrelated to the output variable as well as attributes that are very similar to each other.
- BLR is easy to implement and very efficient to train. it is possible to use BLR as a benchmark and try using more complex algorithms from there on.
- BLR is a good baseline that it can be to use to measure the performance of other more complex algorithms.

○ Disadvantages of BLR

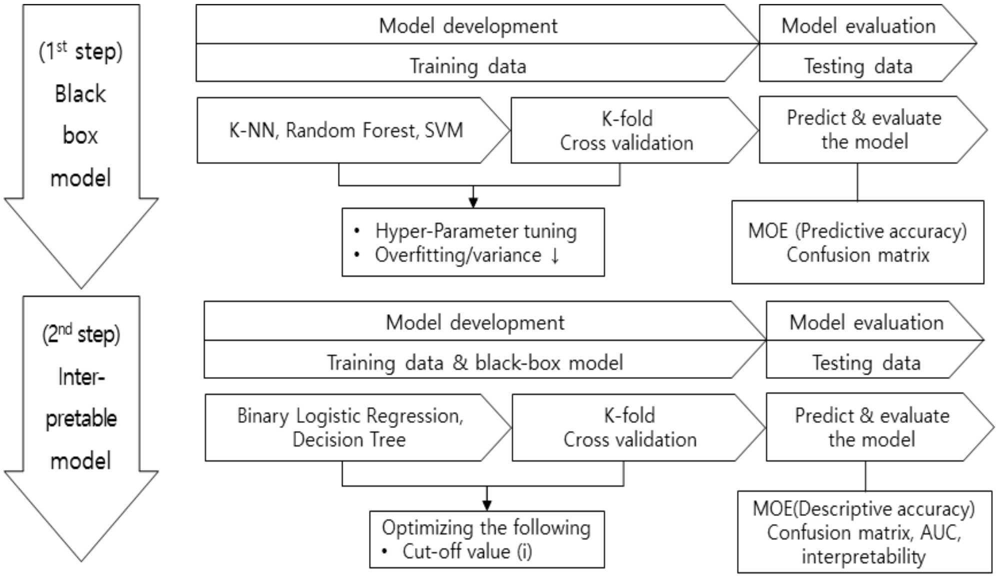
- BLR can't solve non-linear problems since it's decision surface is linear.
- BLR is also not one of the most powerful algorithms out there and can be easily outperformed by more complex ones.
- BLR is not a useful tool unless you have already identified all the important independent variables.
- BLR can only predict a categorical outcome.
- BLR is also an algorithm that is known for its vulnerability to over-fitting.

4.2. Model development

4.2.1. Procedure

In this study, a statistical analysis package, R-studio (version 1.2.1335), was used to develop a model for the casualty crash prediction. Training and test data were divided into 8:2 proportion for the application of machine learning algorithm for development of black-box model and interpretable model. When dividing training and test data, random sampling process was conducted considering whether or not a casualty crash was occurred.

The procedure of model development is shown in [Figure 4-6].



[Figure 4-6] Procedure of model development

As mentioned earlier, the black-box model in the first step applied three machine learning algorithms which were KNN, RF, and SVM, And they were trained using the K-fold cross validation process for hyper-parameters that required tunings in each methodology. The performance of the three machine learning algorithms was evaluated for accuracy, sensitivity, specificity, and accuracy through the conduction matrix. In this way, the prediction performance of the black-box model was evaluated. Among the three machine learning algorithms, the best methodology was chosen as the black-box model.

Next, DT and BLR were applied to the interpretable model. In the interpretable model, the K-fold cross validation process was used to train for hyper-parameter tuning in the same way as the black-box model development. The performance of the two machine learning algorithms was evaluated for accuracy, sensitivity, specificity, and accuracy through the conduction matrix, and a comparative evaluation of AUC was conducted. In this way, the descriptive accuracy of the interpretable model was evaluated and the best methodology was chosen as the interpretable model.

4.2.2. Measures of effectiveness

In the field of machine learning of statistical classification, confusion matrix is a table that is often used to describe the performance of a classification model on a set of test data for which the true values are known. A confusion matrix is a summary of prediction results on a classification problem as shown in <Table 4-1>. The number of correct and incorrect predictions is summarized with count values and broken down by each class. This is the key to the confusion matrix. The confusion matrix shows the ways in which its classification model is confused when it makes predictions. It gives us insight not only into the errors being made by a classifier but more importantly the types of errors that are being made.

<Table 4-1> Confusion matrix for MOE

Confusion Matrix		Prediction Model		MOE
		Positive	Negative	
Reference	Positive	TP (True Positive)	FN (False Negative)	Sensitivity
	Negative	FP (False Positive)	TN (True Negative)	Specificity
MOE		Precision	-	Accuracy

○ Definition of the terms

- Positive (P): Observation is positive.
- Negative (N): Observation is not positive.
- True Positive (TP): Observation is positive, and is predicted to be positive.
- False Negative (FN): Observation is positive, but is predicted negative.
- True Negative (TN): Observation is negative, and is predicted to be negative.
- False Positive (FP): Observation is negative, but is predicted positive.

<Table 4-2> MOE for machine learning

MOE	Formula	MOE	Formula
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	Sensitivity	$\frac{TP}{TP+FN}$
Specificity	$\frac{TN}{TN+FP}$	Precision	$\frac{TP}{TP+FP}$

The formula for accuracy, sensitivity, specificity, and precision, the MOEs of machine learning to be used in this study, are shown in < Table 4-2>. Accuracy assumes equal costs for both kinds of errors. A 99% accuracy can be excellent, good, fair, poor or terrible depending upon the problem. Sensitivity can be defined as the ratio of the total number of correctly classified positive examples divide to the total number of positive examples. High sensitivity indicates the class is correctly recognized. To get the value of precision we divide the total number of correctly classified positive examples by the total number of predicted positive examples. High precision indicates an example labelled as positive is indeed positive.

- High sensitivity, low precision: This means that most of the positive examples are correctly recognized (low FN), but there are a lot of false positives.
- Low sensitivity, high precision: This shows that we miss a lot of positive examples (high FN), but those we predict as positive are indeed positive (low FP).

4.2.3. K-fold cross validation

Cross validation is a re-sampling procedure used to evaluate machine learning models on a limited data sample. The procedure has a single parameter called k that refers to the number of groups that a given data sample is to be split into. As such, the procedure is often called k-fold cross validation. When a specific value for k is chosen, it may be used in place of k in the reference to the model, such as k=10 becoming 10-fold cross validation.

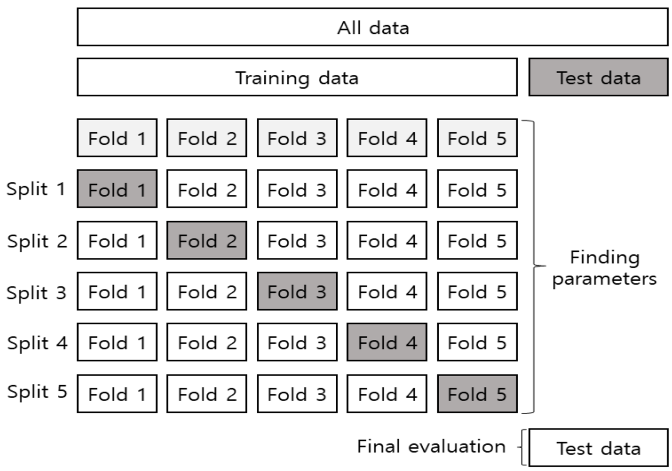
The configuration of k-fold cross validation is shown in [Figure 4-7] and general procedure is as follows.

○ step-1 Partition the original training data set into k equal subsets. Each subset is called a fold. Let the folds be named as $f_{1,}f_2, \cdots, f_k$. For i = 1 to i = k

○ step-2 Keep the fold f_1 as validation set and keep all the remaining k-1 folds in the cross validation training set.

○ step-3 Train your machine learning model using the cross validation training set and calculate the accuracy of your model by validating the predicted results against the validation set.

○ step-4 Estimate the accuracy of your machine learning model by averaging the accuracies derived in all the k cases of cross validation.



[Figure 4-7] K-fold cross validation

Cross validation is primarily used in applied machine learning to estimate the skill of a machine learning model on unseen data. That is, to use a limited sample in order to estimate how the model is expected to perform in general when used to make predictions on data not used during the training of the model. It is a popular method because it is simple to understand and because it generally results in a less biased or less optimistic estimate of the model skill than other methods, such as a simple train/test split.

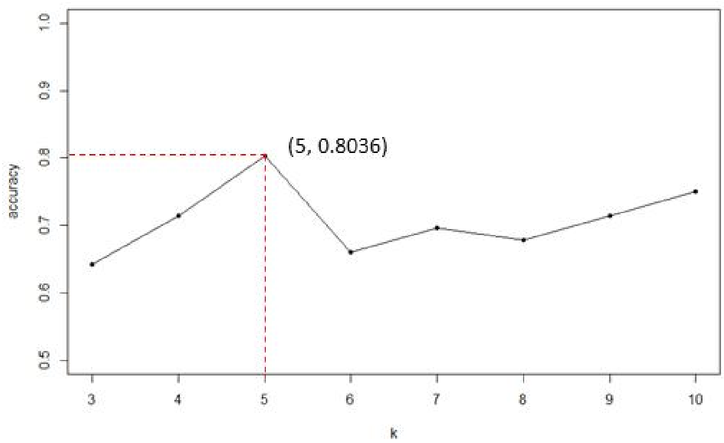
4.3. Result of model development

4.3.1. Result of black-box model

1) KNN

In this study, k values with the best accuracy performance were found through 10-fold cross validation process according to the above KNN algorithm performance procedure.

As the k value increases, the accuracy is also increased and the k value becomes the maximum value when the k value is 5 and then decreases again. The optimal value was found to be 80.36% in case of k=5 as shown in [Figure 4-8].



[Figure 4-8] Validation for optimal k

In addition, k=5 was applied to verify performance of KNN algorithm with the test data, and accuracy was found to be 80.36%.

<Table 4-3> Predicted result of KNN

MOE	Accuracy	Sensitivity	Specificity	Precision
Value	0.8036	0.8889	0.7241	0.7500

The results of evaluating the MOE are shown in <Table 4-3>: accuracy = 0.8036, sensitivity = 0.8889, specificity = 0.7241, and precision = 0.75.

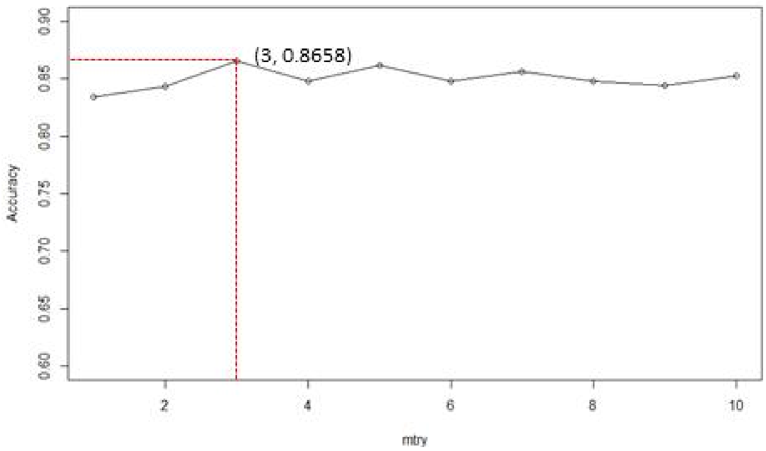
2) RF

There are many hyper-parameters that RF has to set up to find the best performance. The main hyper-parameters which used in this study are following:

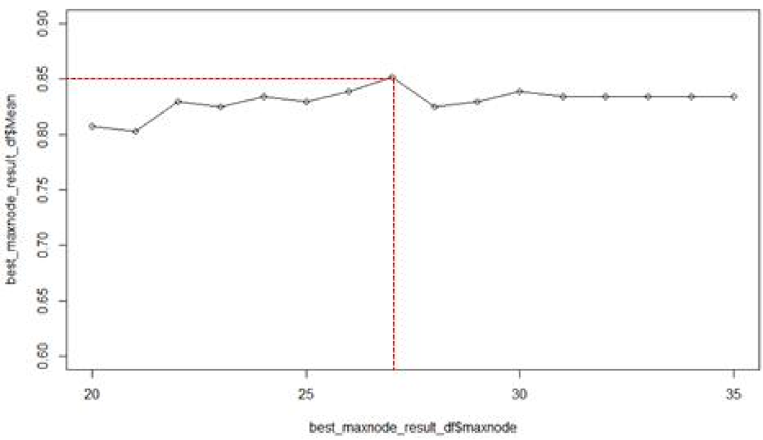
- m-try: Number of variables randomly sampled as candidates at each split. Note that the default values are different for classification ($\sqrt{p}$ where p is number of variables in x) and regression ($\frac{p}{3}$).
- Max-nodes: Maximum number of terminal nodes trees in the forest can have. If not given, trees are grown to the maximum possible (subject to limits by node-size). If set larger than maximum possible, a warning is issued.
- n-tree: Number of trees to grow. This should not be set to

too small a number, to ensure that every input row gets predicted at least a few times.

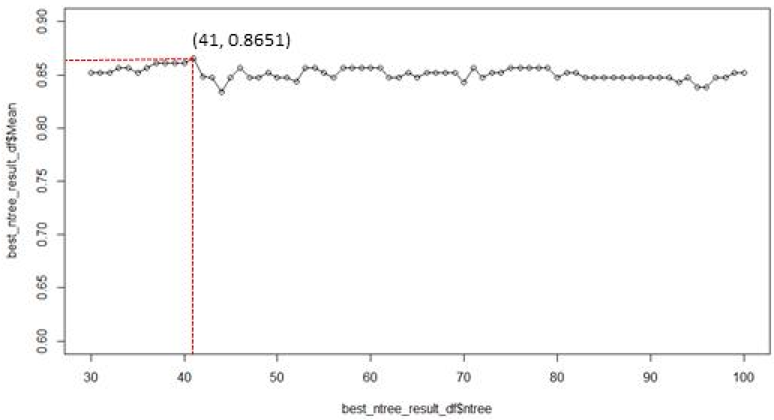
In this study, three hyper-parameters with the best accuracy performance were found through 10-fold cross validation process as shown in [Figure 4-9] to [Figure 4-11].



[Figure 4-9] Tuning the hyper-parameter (m-try)



[Figure 4-10] Tuning the hyper-parameter (max-nodes)



[Figure 4-11] Tuning the hyper-parameter (n-tree)

The resulting "best" hyper-parameters are as follows: m-try = 3, max-nodes = 27 and n-tree = 41. Again, a new RF algorithm was run using these values as hyper-parameter inputs to evaluate the performance through test data, and accuracy was found to be 83.93%.

<Table 4-4> Predicted result of RF

MOE	Accuracy	Sensitivity	Specificity	Precision
Value	0.8393	0.8333	08438	0.8000

The results of evaluating the MOE are in shown <Table 4-4>: accuracy = 0.8393, sensitivity = 0.8333, specificity = 0.8438, and precision = 0.8.

3) SVM

There are many hyper-parameters that SVM has to set up to find the best performance. The main hyper-parameters which used in this study are followings:

- kernel: the kernel type to be used.
  The most common kernels are radial basis function (this is the default value), polynomial or sigmoid, but it is possible to create researcher's own kernel.
- c_0 (cost): it means the SVM optimization how much you want to avoid miss-classifying each training example.
  If the c_0 is higher, the optimization will choose smaller margin hyper-plane, so training data miss-classification rate will be lower.
  If the c_0 is low, then the margin will be big, even if there will be miss-classified training data examples.
- γ (gamma): it defines how far the influence of a single training example reaches.
  This means that high γ will consider only points close to the plausible hyper-plane and low γ will consider points at greater distance.
- d (degree): it is used only if the chosen kernel is poly and sets the degree of the polynomial.

<Table 4-5> Formula and parameters for kernel functions in the SVM

Kernel	Formula	Parameters
Linear	$u \top v$	(none)
Polynomial	$(\gamma u \top v + c_0)d$	γ, d, c_0
Radial basis function	$\exp(-\gamma \lvert u - v \rvert^2)$	γ
Sigmoid	$\tanh(\gamma u \top v + c_0)$	γ, c_0

The kernel functions of used in this study are linear, polynomial, radial basis function and sigmoid. The hyper-parameters that can be tuned for each kernel function are in shown <Table 4-5>.

In this study, hyper-parameters with the best accuracy performance were found through 10-fold cross validation process as shown in <Table 4-6>. The linear kernel does not require parameter tuning, and the polynomial kernel showed the highest accuracy when γ = 1, d = 3, and c_0 =2. In addition, the radial basis function kernel showed the highest accuracy when γ = 1, and the sigmoid kernel showed the highest accuracy when γ = 0.0625 and c_0 = 1.

<Table 4-6> Best parameter for kernel functions in the SVM

Kernel	γ	d	c_0
Linear	-	-	-
Polynomial	1	3	2
Radial basis function	1	-	-
Sigmoid	0.0625	-	1

The results of the performance for accuracy, sensitivity, specificity, and precision for each of the four kernel functions by applying hyper-parameters optimized through the 10-fold cross validation process are shown in <Table 4-7>.

<Table 4-7> Predicted result of SVM

Kernel function	Accuracy	Sensitivity	Specificity	Precision
Linear	0.7321	0.7895	0.7027	0.5769
Polynomial	0.8750	0.8421	0.8919	0.8000
Radial basis function	0.7857	0.5789	0.8919	0.7333
Sigmoid	0.7143	0.8421	0.6486	0.5517

The results of evaluation for accuracy, sensitivity, specificity, and precision for four kernel functions show that polynomial kernel function is the best for all of MOEs (in terms of accuracy, sensitivity, specificity, and precision).

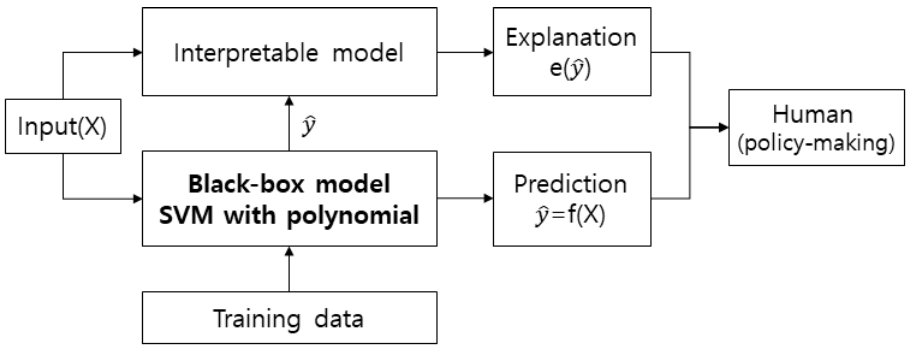
4) Selection of the black-box model

In this study, KNN, RF, and SVM machine learning algorithms were applied as black-box models for IML, and the 10-fold cross validation process optimized hyper-parameters for each machine learning algorithm. Performance evaluations for each model showed that SVMs with a polynomial kernel had the highest performance on all MOEs of accuracy, sensitivity, specificity, and precision as shown in <Table 4-8>.

<Table 4-8> Predicted result of the black-box models

Model	Accuracy	Sensitivity	Specificity	Precision
KNN (k=5)	0.8036	0.8889	0.7241	0.7500
RF	0.8393	0.8333	0.8438	0.8000
SVM (polynomial)	0.8750	0.8421	0.8919	0.8000

Therefore, SVM with polynomial kernel was chosen as the black-box model for developing the casualty crash prediction model as shown in [Figure 4-12].

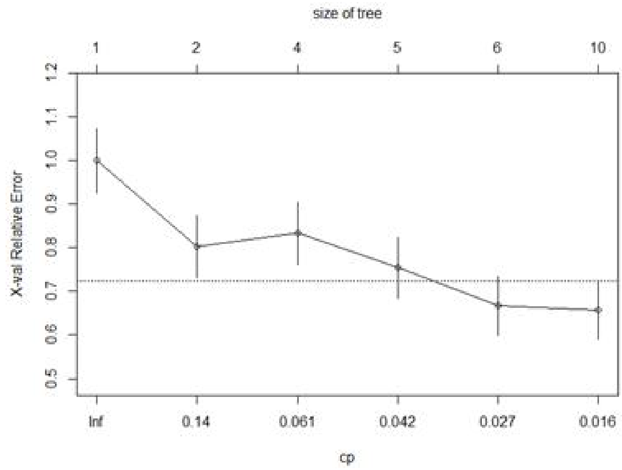


[Figure 4-12] Black-box model selection

4.3.2. Result of interpretable model

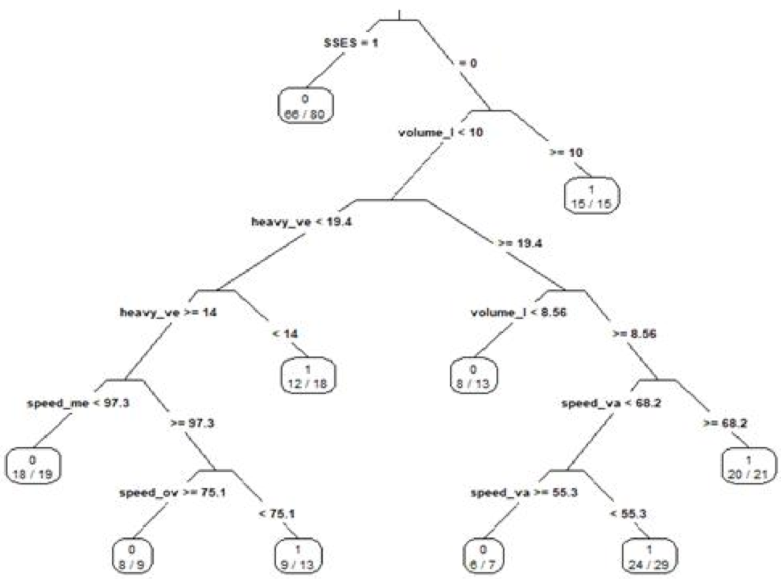
1) DT

It is necessary to optimize the decision tree through the pruning process, because there are concerns about over fitting. It is common to find size of tree that minimizes variances through the 10-fold cross validation. The result of pruning process is shown in [Figure 4-13].



[Figure 4-13] Tuning the size of tree

X-error is minimized when size of tree is 10 as shown in [Figure 4-13], and variables used in tree construction are HVR, SM, SOR, SV, SSES, and TVL as shown in [Figure 4-14].



[Figure 4-14] Result of DT

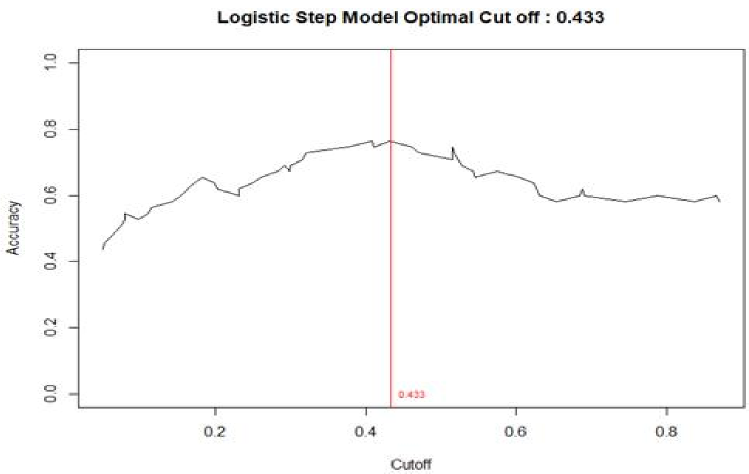
The results of evaluating the MOE are shown in <Table 4-9>: accuracy = 0.6964, sensitivity = 0.7143, specificity = 0.6667, and precision = 0.7813.

<Table 4-9> Result of DT

MOE	Accuracy	Sensitivity	Specificity	Precision
Value	0.6964	0.7143	0.6667	0.7813

2) BLR

Generally, when applying BLR, the cut-off value for binary classification is applied as 0.5. In this study, when developing a model that used BLR to predict the occurrence of casualty crash, the cut-off value which had the highest accuracy performance was found through 10-fold cross validation process and it was 0.433 as shown in [Figure 4-15].



[Figure 4-15] Tuning the optimal cut-off value

The results of evaluating the MOE are shown in <Table 4-10>: accuracy = 0.7636, sensitivity = 0.7917, specificity = 0.7419, and precision = 0.7073.

<Table 4-10> Result of BLR

MOE	Accuracy	Sensitivity	Specificity	Precision
Value	0.7636	0.7917	0.7419	0.7073

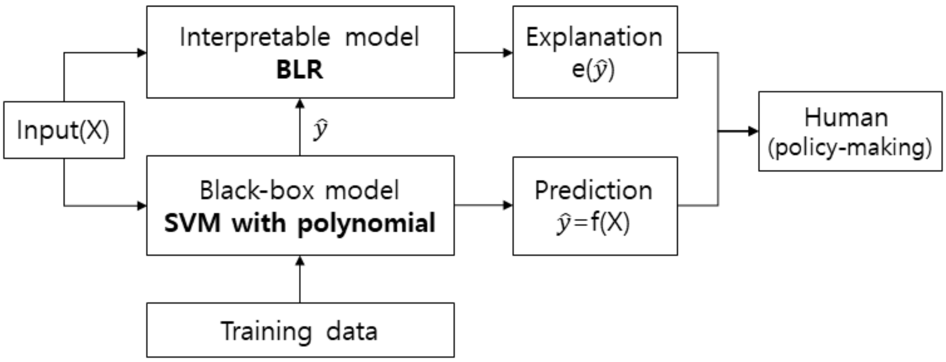
3) Selection of the interpretable model

In this study, DT and BLR machine learning algorithms were applied as interpretable models for IML and the 10-fold cross validation process optimized hyper-parameters for each machine learning algorithm. Comparing the performance of DT to BLR, the sensitivity and specificity of BLR was higher than that of DT, whereas the precision of DT were higher than that of BLR. Because the BLR is higher than the DT for total accuracy, BLR was selected as an interpretable model. The predicted result of comparison between interpretable models is shown in <Table 4-11>.

<Table 4-11> Predicted result of the interpretable models

Model	Accuracy	Sensitivity	Specificity	Precision
DT	0.6964	0.7143	0.6667	0.7813
BLR	0.7636	0.7917	0.7419	0.7073

For the development of IML models, SVM with polynomial kernel was applied as black-box model to increase the predictive accuracy and BLR was applied as an interpretable model to increase the descriptive accuracy. The final result of IML development is shown in [Figure 4-16].



[Figure 4-16] Result of model development for IML

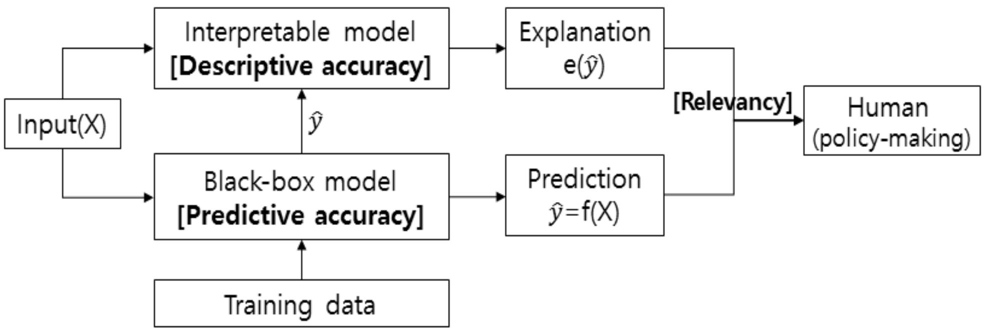
In the chapter 5, the performance evaluation will be performed by comparing with the typical BLR model based on the above IML model development results.

5. Evaluation & Application

5.1. Evaluation

5.1.1. The PDR framework for IML

In general, it is unclear how to select and evaluate interpretation methods for a particular problem. To help guide this process, Murdoch et al. (2018) introduced the PDR framework, consisting of three desiderata that should be used to select interpretation methods for a particular problem: predictive accuracy, descriptive accuracy, and relevancy. The configuration of PDR framework for IML is shown in [Figure 5-1].



[Figure 5-1] PDR framework for IML

The information produced by an interpretation method should be faithful to the underlying process the practitioner is trying to understand. In the context of machine learning, there are two areas where errors can arise: when approximating the underlying data relationships with a model (predictive accuracy) and when approximating what the model has learned using an interpretation method (descriptive accuracy). For an interpretation to be trustworthy, one should try to maximize both of the accuracies. Evaluating the quality of a model's fit has been well studied in supervised machine learning frameworks, through measures such as test-set accuracy. In the context of interpretation, this error is described as predictive accuracy. It is possible to define descriptive accuracy, in the context of interpretation, as the degree to which an interpretation method objectively captures the relationships learned by machine learning models. In selecting what model to use, practitioners are often faced with a trade-off between predictive and descriptive accuracy. The simplicity of model-based interpretation methods yields consistently high descriptive accuracy, but can sometimes result in lower predictive accuracy on complex data-sets. On the other hand, in complex settings such as image analysis, complicated models generally provide high predictive accuracy, but are harder to analyze, resulting in a lower descriptive accuracy.

It is possible to define an interpretation to be relevant if it provides insight for a particular audience into a chosen domain problem. Relevancy often plays a key role in determining the trade-off between predictive and descriptive accuracy. Depending on the context of the problem at hand, a practitioner may choose to focus on one over the other. For instance, when interpretability is used to audit a model's predictions, such as to enforce fairness, descriptive accuracy can be more important. In contrast, interpretability can also be used solely as a tool to increase the predictive accuracy of a model, for instance, through improved feature engineering.

5.1.2. Predictive accuracy

Evaluating the quality of a model's fit has been well studied in supervised machine learning frameworks, through measures such as test-set accuracy. In the context of interpretation, this error is described as predictive accuracy. This is used to evaluate the prediction performance of IML's black-box model. This study compared the predicted performance of the black-box model IML-SVM with the typical BLR model and evaluated the predictive accuracy. The predicted result of comparison between BLR and IML-SVM is shown in <Table 5-1>.

<Table 5-1> Predictive accuracy

Model	Accuracy	Sensitivity	Specificity	Precision
BLR	0.6545	0.7500	0.5806	0.5806
IML-SVM (Black-box model)	0.8750	0.8421	0.8919	0.8000

Comparing the predictive accuracy of IML-SVM to BLR, the IML-SVM is higher than the BLR for all of MOEs which are accuracy, sensitivity, specificity, and precision. In particular, for total accuracy, the IML-SVM, which is applied as a black-box model, outperformed the BLR by about 22%.

5.1.3. Descriptive accuracy

It is possible to define descriptive accuracy, in the context of interpretation, as the degree to which an interpretation method objectively captures the relationships learned by machine learning models. This is used to evaluate the prediction performance of IML's interpretable model. This study compared the predicted performance of the interpretable model IML-BLR with the typical BLR model and evaluated the descriptive accuracy. The predicted result of comparison between BLR and IML-BLR is shown in <Table 5-2>.

<Table 5-2> Descriptive accuracy

Model	Accuracy	Sensitivity	Specificity	Precision
BLR	0.6545	0.7500	0.5806	0.5806
IML-BLR (Interpretable model)	0.7636	0.7917	0.7419	0.7073

Comparing the descriptive accuracy of IML-BLR to BLR, the IML-BLR is higher than the BLR for all of MOEs which are accuracy, sensitivity, specificity, and precision. In particular, for total accuracy, the IML-BLR, which is applied as a interpretable model, outperformed the BLR by about 10%.

In case of the BLR, specificity which value is 0.5806 is very low. It means that there are many cases in which casualty crashes are predicted to occur in sections where no actual accidents have occurred. Thus, it may face the criticism for over-investment in traffic safety facilities in such cases.

<Table 5-3> Confusion matrix of IML-BLR

Confusion Matrix		Predicted value	
		Positive ('1')	Negative ('0')
Actual value	Positive ('1')	19	5
	Negative ('0')	8	23

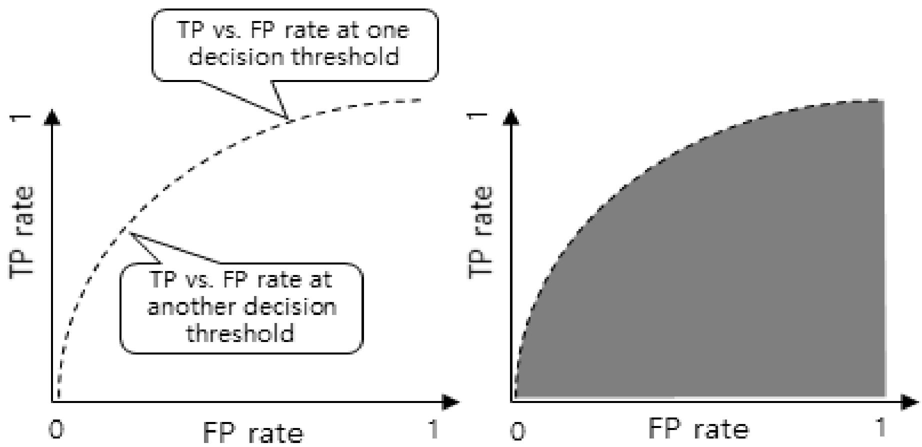
<Table 5-3> shows the conduction matrix for the IML-BLR predicted results for the test data. Of the total test data, there are 5 cases of FN (False Negative) and 8 cases of FP (False Positive). The sample raw data for theses are shown in <Table 5-4>.

<Table 5-4> List of miss-classification

Actual	Predicted	Probability	SSES	SOR	SV	ln(TVL)	HVR	CR
1	0	0.079923	1	0	36.98	9.38	15.8	3.28
1	0	0.183197	1	0	36.29	8.59	22.5	31.04
1	0	0.198303	0	7.22	26.87	8.45	19.57	6.36
1	0	0.202937	1	19.25	58.18	8.28	18.33	39.11
1	0	0.292531	0	17.20	26.93	8.68	18.22	14.23
0	1	0.440532	0	12.52	62.06	7.49	35.23	4.44
0	1	0.515073	0	64.6	88.25	8.53	17.7	7.89
0	1	0.515534	0	17.09	42.53	8.87	25.73	13.25
0	1	0.545896	1	14.11	33.26	10.24	32.79	3.01
0	1	0.653620	0	88.84	49.77	8.02	14.26	39.11
0	1	0.683623	0	0	149.16	9.94	16.85	12.35
0	1	0.745368	0	17.74	64.7	9.66	29.26	10.32
0	1	0.837873	0	99.73	4.52	9.55	24.27	3.43
Average of all samples				27.08	67.74	9.15	20.04	11.59

In three of the five FN (False Negative) cases, there was a low probability of casualty crash occurrence because of SSES installation, but casualty crashes occurred in reality. This is due to other independent variables that cannot be explained by the crash prediction model developed in this study, therefore further studies are needed. On the other hand, in the case of FP (False Positive), there was a high probability of casualty crash occurrence, but casualty crashes did not occur in reality. Therefore an additional analysis for other safety conditions is also needed.

Next, a comparative evaluation of AUC (Area Under the ROC Curve) - ROC (Receiver Operating Characteristics) curves between BLR and IML-BLR was conducted. AUC - ROC are used a lot in addition to the confusion matrix when evaluating the results of machine learning.



[Figure 5-2] ROC and AUC

It is one of the most important evaluation metrics for checking any classification model's performance. It is also written as AUROC. AUROC curve is a performance measurement for classification problem at various thresholds settings. ROC is a probability curve and AUC represents degree or measure of separability. It tells how much model is capable of distinguishing between classes. Higher the AUC, better the model is at predicting 0s as 0s and 1s as 1s.

The ROC curve is plotted with TPR (True Positive Rate) against the FPR (False Positive Rate) where TPR is on y-axis and FPR is on the x-axis as shown in [Figure 5-2].

- TPR : True Positive Rate (=sensitivity)

$$TPR = \frac{TP}{TP + FN}$$

- FPR : False Positive Rate (=1-specificity)

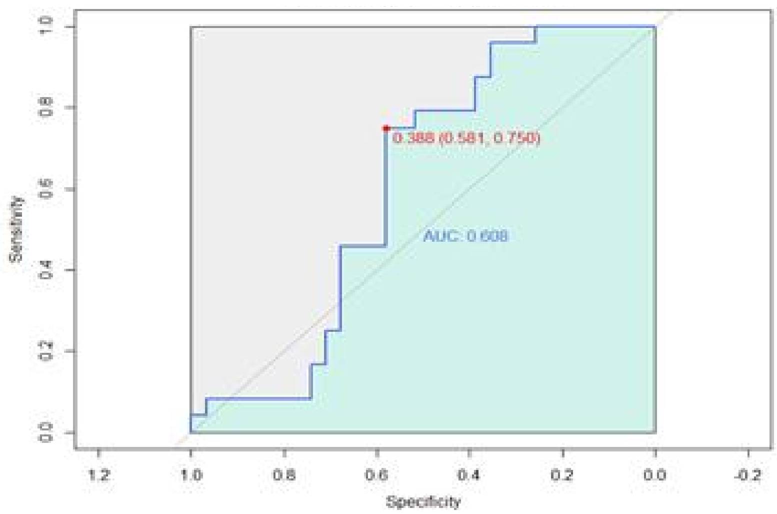
$$FPR = \frac{FP}{FP + TN}$$

An excellent model has AUC near to the 1 which means it has good measure of separability. A poor model has AUC near to the 0 which means it has worst measure of separability. In fact it means it is reciprocating the result. It is predicting 0s as 1s and 1s as 0s. And when AUC is 0.5, it means model has no class separation capacity whatsoever.

Model performance according to AUC value can be following:

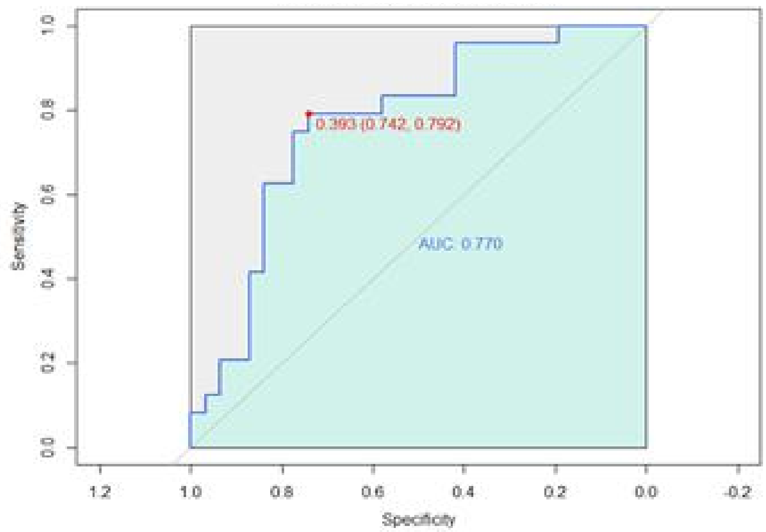
- excellent = 0.9 ~ 1.0
- good = 0.8 ~ 0.9
- fair = 0.7 ~ 0.8
- poor = 0.6 ~ 0.7
- fail = 0.5 ~ 0.6

In this study, the AUROC curves of BLR and IML-BLR were compared for evaluation of performance on the descriptive accuracy of the interpretable model. AUROC curve of BLR is shown in [Figure 5-3] and that of IML-BLR is shown in [Figure 5-4].



[Figure 5-3] AUROC curve of BLR

The AUC value of the BLR was 0.608 and it means that the performance of the model is poor (0.6 ~ 0.7). Whereas the AUC value of the IML-BLR was 0.770 which was 0.162 higher than the AUC value of the BLR. It means that the performance of the model is fair (0.7 ~ 0.8).



[Figure 5-4] AUROC curve of IML-BLR

In general, crash prediction models do not often have high performance in the model because crashes occur very randomly. This study also shows that predictive performance of typical BLR can be improved to fair level through the IML methodology.

5.1.4. Relevancy

Relevancy can be defined as an interpretation to be relevant if it provides insight for a particular audience into a chosen domain problem. In other words, it is the ability to explain or to present in understandable terms to a human (Doshi-Velez, 2019). In this study, to compare IML-BLR and BLR from a relevancy point of view, the independent variables applied to the casualty crash prediction model developed by IML were applied equally to the BLR to compare the coefficients and significant probabilities of the estimated independent variables. The following <Table 5-5> and <Table 5-6> show coefficient, estimated value and significant probabilities for the result of model development.

<Table 5-5> Result of IML-BLR

Coefficient	Estimate	Std. Error	Z-value	Pr (>|z|)
(Intercept)	-10.509254	2.631240	-3.994	6.5e-05***
SSES	-1.210235	0.398248	-3.039	0.002374***
Speed_Over_Ratio (SOR)	0.020798	0.005764	3.608	0.000308***
Speed_Variance (SV)	0.005467	0.002816	2.542	0.042195**
ln(Traffic_Volume_Lane) (TVL)	0.788860	0.258576	3.051	0.002282***
Heavy_Vehicle_Ratio (HVR)	0.099162	0.030133	3.291	0.000999***
Curve_Ratio (CR)	0.032822	0.019088	2.719	0.045525**

**p<0.05, ***p<0.01

First of all, for IML-BLR as shown in <Table 5-5>, SSES, SOR, SV, TVL, HVR, and CR variables were selected as independent variables. For the sign of variables, the installation of SSES resulted in fewer casualty crashes, and for SV, SOR, TVL, and CR, it was shown that the increase in its size resulted in more casualty crashes. The sign of the all independent variables can all be seen as appropriate from a human point of view. In addition, the significance probability for all independent variables was shown to be statistically significant at the least 95% confidence level.

<Table 5-6> Result of BLR

Coefficient	Estimate	Std. Error	Z-value	Pr (>|z|)
(Intercept)	-14.697326	2.989560	-4.916	8.82e-07***
SSES	-1.648814	0.430585	-3.829	0.000129***
Speed_Over_Ratio (SOR)	0.022930	0.006159	3.723	0.000197***
Speed_Variance (SV)	0.004782	0.003175	1.506	0.332024
Traffic_Volume_Lane (ln(TVL))	1.278155	0.301404	4.241	2.23e-05***
Heavy_Vehicle_Ratio (HVR)	0.091724	0.031832	2.882	0.003958***
Curve_Ratio (CR)	0.028513	0.019634	1.452	0.146445

**p<0.05, ***p<0.01

Next, for BLR as shown in <Table 5-6>, the selected independent variables were the same as IML-BLR and their signs were the same. But, significant probability for SV and CR variables was found to be not statistically significant at the 95% confidence level.

When the two models from a human understanding point of view are compared and analyzed, the results of the BLR can be judged to be inappropriate to apply due to the very low specificity. If specificity is low, the criticism of over-investment can be occurred in traffic safety facilities. On the other hand, IML-BLR's results are appropriate at the sign and significance probability levels of the variables applied to model development, and the difference between sensitivity and specificity is not large, so IML-BLR is appropriate in terms of utilization of development results.

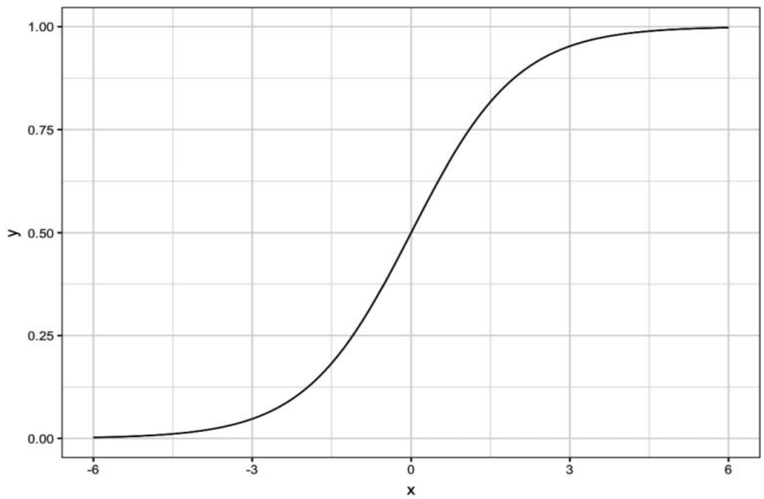
5.2. Impact of Casualty Crash Reduction

5.2.1. Quantification of the effectiveness

In this study, BLR function is used to quantify the effect of SSES installation. Instead of fitting a straight line or hyper-plane, the BLR model uses a non-linear function, the BLR to squeeze the output of a linear equation between 0 and 1. The BLR function is defined as:

$$\log istic(\eta) = \frac{1}{1+\exp(-\eta)}$$

And it is shown in [Figure 5-5].



[Figure 5-5] Binary logistic function

The step from linear regression models to BLR is kind of straightforward. For the classification we prefer probabilities, which are between 0 and 1, so we wrap the right side of the equation into the BLR function and like that force the output to only take on values between 0 and 1.

$$P(y_i = 1) = \frac{1}{1 + \exp(-(\beta_0 + \beta_0 x_{i,1} + \ldots + \beta_p x_{i,p}))}$$

Therefore, the probability formular of casualty crash occurrence based on the IML development result is as follows:

$$P(casaulty = 1)$$
$$= \frac{1}{1 + e^{-(-10.5093 - 1.2102SSES + 0.0208SOR + 0.0055SV + 0.7889TVL + 0.0992HVR + 0.0328CR)}}$$

The installation effect of SSES was quantified using the probability of casualty crash occurrence according to the following steps:

○ step-1 The values for the other independent variables except SSES are replaced by the average values in the probability of casualty crash equation.

○ step-2 Set the SSES value to zero and calculate the probability of casualty crash occurrence before installation.

○ step-3 Set the SSES value to 1 and calculate the probability of casualty crash occurrence after installation.

○ step-4 Probability differences before-after installation quantify the installation effects of SSES.

According to the above procedure, the probability of casualty crash occurrence before SSES installation is 51% and the probability of casualty crash occurrence after SSES installation is 23%. Therefore, it is possible to confirm that SSES installation reduces the probability of casualty crash occurrence by about 28%.

In addition, the probability of casualty crash occurrence after the installation of SSES can be verified through the probability equation of the developed IML model. The estimated coefficients of SSES can be expressed as follows:

$$Odds\,ratio = \frac{p(Y=1)}{1-p(Y=1)} = e^{-1.210235} = 0.297$$

$$\therefore p(Y=1) = 0.23$$

In other words, the probability ($p(Y=1)$) of casualty crash occurrence in case of an SSES installation will be about 0.297 times lower than in case of no installation. In addition, the probability of casualty crash occurrence after SSES installation is calculated based on the above equation as 23%. It can be confirmed that this result is the same as the probability of casualty crash occurrence after installation of SSES.

The results of quantifying the SSES installation effects derived from this study were compared with the results of the relevant prior studies. The effects of SSES installation in the preceding studies are shown in <Table 5-7>. The prior studies were classified as foreign and domestic cases. And the effects of accident reduction were divided into total crashes and casualty crashes. If SSES is installed, it can be confirmed that the total crashes have a reduction effect of about 22 to 50%, and that casualty crashes have a reduction effect of about 18 to 42%. The 28% of probability of reducing casualty crashes quantified through this study is within the range of the reduction in casualty crash in the prior studies. However, in the case of prior studies, it is about the effect of decreasing the number of casualty crashes, it is difficult to make a direct comparison as it is the effect of decreasing the probability of casualty crash occurrence.

<Table 5-7> Effectiveness of crash reduction in literature reviews

	Author &Subject	Sites	Effectiveness
Foreign	• Torre et al. (2019), safety effects of automated section speed control on the Italian motorway network	125	Total crash: 22%↓ Fatal injury 18%↓
	• Montella et al. (2015), Effect on speed and safety of point-to-point speed enforcement systems	1	Total crash: 32%↓ Injury crash: 37%↓
Domestic	• Jung et al. (2014), Traffic accident reduction effects of Section Speed Enforcement System(SSES) Operation in Freeways	9	Total crash: 32%↓ Fatal injury: 42%↓
	• Yun et al. (2011), Effectiveness of the point-to-point speed enforcement system	8	Total crash: 50%↓
	• Lee et al. (2013), A Study on the Analysis for the Effects of the Section Speed - Enforcement System at the Misiryeong tunnel section	1	Total crash: 46%↓

The effect of reducing the probability of casualty crash occurrence to be derived from this study can not be ascertained by the prior studies, rather, is the basis for confirming the differentiation of this study methodology.

5.2.2. Mediation effect analysis

1) Mediation effect

In statistics, a mediation model seeks to identify and explain the mechanism or process that underlies an observed relationship between an independent variable and a response variable via the inclusion of a third hypothetical variable, known as a mediator variable. Rather than a direct relationship between the independent variable and the response variable, a mediation model proposes that the independent variable influences the mediator variable, which in turn influences the response variable. Thus, the mediator variable serves to clarify the nature of the relationship between the independent and response variables.

Mediation analysis is employed to understand a known relationship by exploring the underlying mechanism or process by which one variable influences another variable through a mediator variable. Mediation analysis facilitates a better understanding of the relationship between the independent and response variables, when the variables appear to not have a definite connection. They are studied by means of operational definitions and have no existence apart.

The basic conceptual framework of a mediation process with a single mediator is shown in [Figure 5-7]. Treatment (T) can impact the outcome (Y) either indirectly via the mediator (M) or directly. In health management interventions we may expect a significant proportion of the effect to be direct, since there are likely to be myriad variables not observed through the mediated pathway (including other unmeasured mediators). Thus, the total treatment effect is the sum of both direct and indirect effects. These associations can be expressed statistically using the following set of linear regressions:

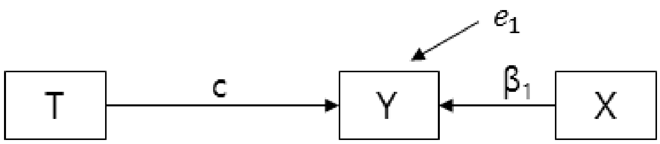
$$Y = i_1 + cT + \beta_1 X + e_1$$

$$M = i_2 + aT + \beta_2 X + e_2$$

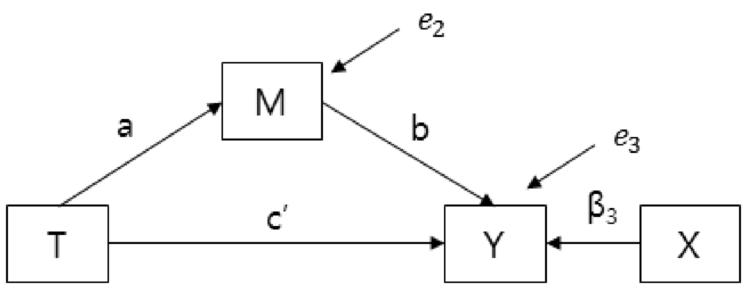
$$Y = i_1 + c'T + bM + hTM + \beta_3 X + e_3$$

- c : the total effect of T on Y
- ab : indirect effect of T on Y
- c' : the direct effect of T on Y after controlling for M
- $c' = c - ab$
- h : interaction effect of T and M on Y
- T : treatment
- M : mediator
- X : covariates
- Y : outcome
- i_1, i_2, i_3 : intercepts
- e_1, e_2, e_3 : unexplained or error variance

First equation is a standard outcomes model estimating the average total effect of the intervention by regressing the outcome Y on the treatment variable T and one or more pre-intervention characteristics X. Second equation represents the a pathway in [Figure 5-7] in which the mediator M is regressed on T and X. Third equation provides both the b and c' pathways indicated in [Figure 5-7] by regressing the outcome on T, M, and X.



[Figure 5-6] Total effect model



[Figure 5-7] Mediation effect model

In the [Figure 5-6] and [Figure 5-7] as shown, the indirect effect is the product of path coefficients "a" and "b". The direct effect is the coefficient "c'". The direct effect measures the extent to which the response variable changes when the independent variable increases by one unit and the mediator variable remains unaltered. In contrast, the indirect effect measures the extent to which the response variable changes when the independent variable is held fixed and the mediator variable changes by the amount it would have changed had the independent variable increased by one unit.

The indirect effect constitutes the extent to which the T variable influences the Y variable through the mediator. In linear systems, the total effect is equal to the sum of the direct and indirect (c′ + ab in [Figure 5-7] as shown). Whereas in non-linear models, the total effect is not generally equal to the sum of the direct and indirect effects, but to a modified combination of the two.

2) Mediation analysis

The methods implemented via mediation rely on the following identification result obtained under the sequential ignorability assumption of Imai et al. (2010).

$$\bar{\delta}(t) = \iint E(Y_i|M_i = x, T_i = t, X_i = x)\, dF_{M_i|T_i=1,X_i=x}(m) - dF_{M_i|T_i=0,X_i=x}(m)\, dF_{X_i}(x)$$

$$\bar{\zeta}(t) = \iint E(Y_i|M_i = m, T_i = 1, X_i = x) - E(Y_i|M_i = m, T_i = 0, X_i = x)\, dF_{M_i|T_i=t,X_i=x}(m)\, dF_{X_i}(x)$$

- $\bar{\delta}(t)$: the average mediation (indirect) effect
- $\bar{\zeta}(t)$: the average direct effect
- Y_i, M_i, T_i, X_i : the observed outcome, mediator, treatment, and pre-treatment covariates. respectively

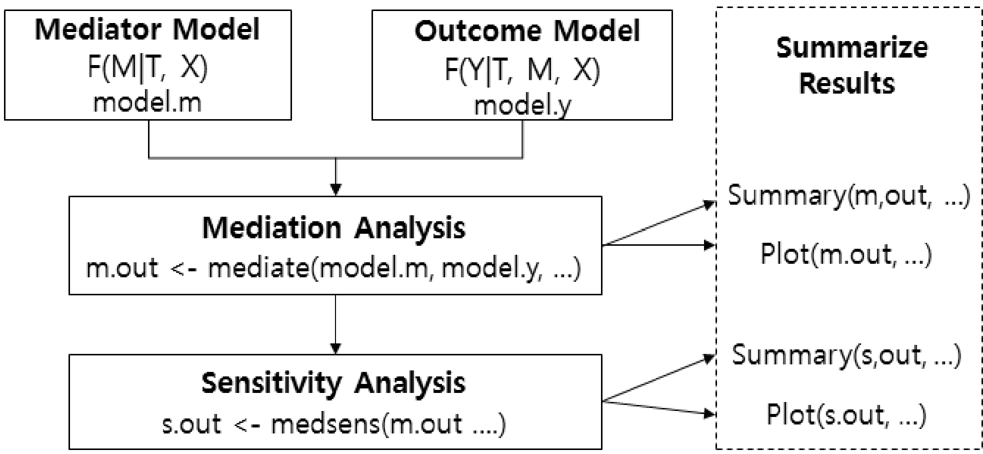
The sequential ignorability assumption states that the observed mediator status is as if randomly assigned conditional on the randomized treatment variable and the pre-treatment covariates.

Mediation analysis under this assumption requires two statistical models;

- the mediator model: $f(M_i|T_i, X_i)$

- the outcome model: $f(Y_i|M_i, T_i, X_i)$

Once these models are chosen and fitted by researchers, then mediation will compute the estimated mediation and other relevant estimates using the algorithms proposed in Imai et al. (2010). The algorithms also produce uncertainty estimates such as standard errors and confidence intervals, based on either a non-parametric bootstrap procedure or a quasi-Bayesian Monte Carlo approximation.



[Figure 5-8] Structure of the mediation package as of version 4.0

In this study, mediation packages are used for mediation effect analysis and its structure is as shown in [Figure 5-8]. The first step is to fit the mediator and outcome models using, for example, regression models with the usual [lm] or [glm] functions. In the second step, the analysts takes the output objects from these models, which in [Figure 5-8] we call [model.m] and [model.y], and use them as inputs for the main function [mediate]. This function then estimates the causal mediation (indirect) effect, direct effect, and total effect along with their uncertainty estimates. Finally, sensitivity analysis can be conducted via the function [medsens] which takes the output of [mediate] as an input. For these outputs, there are both [summary] and [plot] methods to display numerical and graphical summaries of the analyses, respectively.

<Table 5-8> Type of models possible to estimate mediation effects

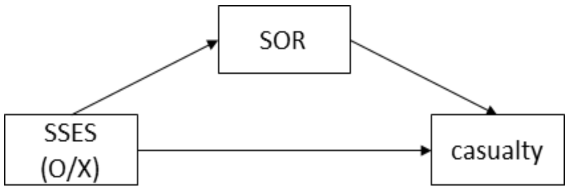
Mediator Model Types	Outcome Model Types				
	Linear	GLM	Ordered	Censored	Semi-parametric
Linear	○	○	○*	○	○*
GLM (BLR)	○	○	○*	○	○*
Ordered	○	○	○*	○	○*
Censored(tobit)	-	-	-	-	-
semi-parametric	○*	○*	○*	○*	○*

* indicate the model combinations that can only be estimated using the non-parametric bootstrap

The mediation packages make it possible to estimate mediation effects as shown in <Table 5-8>. In this study, mediator model type is linear and outcome model type is GLM (e.g. BLR).

3) Mediation effect of SOR

The effect of SSE installation (treatment) on the casualty crash (outcome) through speed-over-ratio (mediator) was analyzed as shown in [Figure 5-9].



[Figure 5-9] Mediation effect model of SOR

For this, mediation analysis under this assumption requires two statistical models;

- the mediator model: $f(M_i|T_i, X_i)$

$$SOR = i_2 + aSSES + \beta_1 SV + \beta_2 TVL + \beta_4 HVR + \beta_4 CR + e_2$$

- the outcome model: $f(Y_i|M_i, T_i, X_i)$

$$P(casualty = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 SSES + \beta_2 SOR + \beta_3 SV + \beta_4 TVL + \beta_4 HVR + \beta_5 CR + e_3)}}$$

- T_i: SSES
- M_i: SOR
- X_i: SV, HVR, TVL, CR
- Y_i: casualty

In <Table 5-9>, ACME (control) is the mediation effect under the control condition, while ACME (treated) is the mediation effect under the treatment condition. The same notation applies to the direct effects.

In this study, to confirm the interaction term between the treatment and the mediator, the method was analyzed by estimating the mediation effect by dividing it into treatment group and control group, and estimating the total effects by the average of each case. Even though the outcome model does not include an interaction term between the treatment and mediator, the estimated effects slightly differ between the treatment and control conditions. This difference, however, is solely due to the non-linearity in the outcome model and should be small.

<Table 5-9> Mediation effect analysis of SOR

Coefficient	Estimate	95% CI Lower	95% CI Upper	p-value
ACME (control)	-0.1249	-0.1866	-0.06	$<2e-16^{***}$
ACME (treated)	-0.0961	-0.1583	-0.04	$<2e-16^{***}$
ADE (control)	-0.2728	-0.3993	-0.13	$<2e-16^{***}$
ADE (treated)	-0.2440	-0.3647	-0.12	$<2e-16^{***}$
Total Effect	-0.3689	-0.4761	-0.26	$<2e-16^{***}$
Prop. Mediated (control)	0.3386	0.1598	0.57	$<2e-16^{***}$
Prop. Mediated (treated)	0.2605	0.1018	0.51	$<2e-16^{***}$
ACME (average)	-0.1105	-0.1703	-0.05	$<2e-16^{***}$
ADE (average)	-0.2584	-0.3816	-0.13	$<2e-16^{***}$
Prop. Mediated (average)	0.2996	0.1324	0.54	$<2e-16^{***}$

**p<0.05, ***p<0.01

* ACME: estimated Average Casual Mediation Effect
* ADE: estimated Average Direct Effect
* When the outcome model is non-linear, the ACME and ADE effect estimates will differ between the treatment and control conditions.

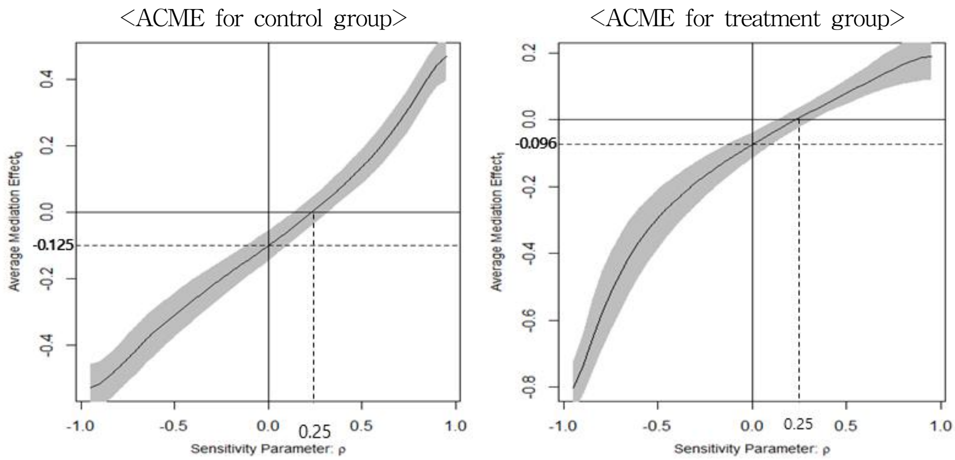
The total effect was estimated to be －0.3689, mediation effect was estimated to be －0.1105 and direct effect was estimated to be −0.2584. The proportion of total effect via mediation was 29.96% as shown in <Table 5-9>.

The causal mediation analysis relies on the sequential ignorability assumption that cannot be directly verified with the observed data. The assumption implies that the treatment is ignorable given the observed pre-treatment confounders and that the mediator is ignorable given the observed treatment and the observed pre-treatment covariates. In order to probe the plausibility of such a key identification assumption, analysts must perform a sensitivity analysis.

Sensitivity analysis is interpreted in terms of a range, and has a high degree of subjectivity, but it may be useful in assessing the degree to which the bias due to the inclusion of confounders may affect the interpretation of the effects. It shows how much the indirect effect changes as a function of ρ (sensitivity parameter) and ρ means the correlation between the error terms of the mediator model and the outcome model. It can be expressed as following;

$$\rho \equiv Corr(e_{i2}, e_{i3}), \quad where\ -1 < \rho < 1$$

If there exist unobserved pre-treatment confounders which affect both the mediator and the outcome, we expect that the sequential ignorability assumption is violated and ρ is no longer zero. The sensitivity analysis is conducted by varying the value of ρ and examining how the estimated ACME changes.

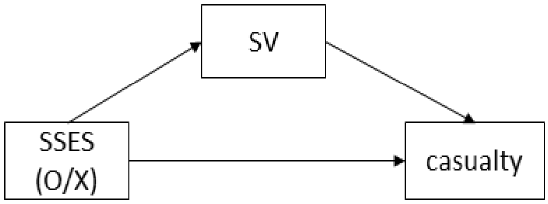


[Figure 5-10] sensitivity analysis of SOR

[Figure 5-11] is a plot of sensitivity analysis and shows, together with the axes of indirect effect and ρ, the observed mediating effect (dashed line) and the values that the indirect effect would reach varying the sensitivity parameter (solid curved line). The confidence interval is represented with a grey background. It can be confirmed that the indirect effect to be zero when ρ is 0.25. It is indicated that the direction of ACME would be maintained unless ρ is more than 0.25.

4) Mediation effect of SV

The effect of SSES installation (treatment) on the casualty crash (outcome) through speed variance (mediator) was analyzed as shown in [Figure 5-12].



[Figure 5-11] Mediation effect model of SV

Similar to the case of mediation analysis for SOR, two statistical models are followings:

- the mediator model: $f(M_i|T_i, X_i)$

$$SV = i_2 + aSSES + \beta_1 SOR + \beta_2 TVL + \beta_4 HVR + \beta_4 CR + e_2$$

- the outcome model: $f(Y_i|M_i, T_i, X_i)$

$$P(casualty=1) = \frac{1}{1+e^{-(\beta_0 + \beta_1 SSES + \beta_2 SOR + \beta_3 SV + \beta_4 TVL + \beta_4 HVR + \beta_5 CR + e_3)}}$$

- T_i: SSES
- M_i: SV
- X_i: SOR, HVR, TVL, CR
- Y_i: casualty

The estimated average mediation effect along with the quasi-Bayesian confidence interval are shown in <Table 5-10>. The total effect was estimated to be -0.302, mediation effect was estimated to be -0.037 and direct effect was estimated to be -0.265. The proportion of total effect via mediation was 12.3%. But it was founded that indirect effects (average ACME=-0.037, p=0.08) and a mediated proportion of 0.123 (p=0.08) is not significant at 95% confidence level. The confidence interval also includes zero for the indirect effect.

<Table 5-10> Mediation analysis of SV

Coefficient	Estimate	95% CI Lower	95% CI Upper	p-value
ACME (control)	-0.041	-0.092	0.00	0.08
ACME (treated)	-0.033	-0.076	0.00	0.08
ADE (control)	-0.269	-0.408	-0.13	<2e-16***
ADE (treated)	-0.260	-0.394	-0.13	<2e-16***
Total Effect	-0.302	-0.428	-0.18	<2e-16***
Prop. Mediated (control)	0.137	0.014	0.33	0.08
Prop. Mediated (treated)	0.108	0.009	0.30	0.08
ACME (average)	-0.037	-0.083	0.00	0.08
ADE (average)	-0.265	-0.399	-0.13	<2e-16***
Prop. Mediated (average)	0.123	0.011	0.32	0.08

**p<0.05, ***p<0.01

* ACME: estimated Average Casual Mediation Effect
* ADE: estimated Average Direct Effect
* When the outcome model is non-linear, the ACME and ADE effect estimates will differ between the treatment and control conditions.

This result is expected as a consequence of imposing a zero correlation between the error terms of the mediator model and the outcome model. Therefore, the sensitivity analysis was not conducted.

5.3. Application for the Korean expressway

In this study, the result of IML model development was validated for applicability in the actual expressway. A prediction model of probability of casualty crash occurrence considering whether or not SSES installation developed through this study was applied to the selection of hazardous sections for Yeongdong Expressway in Korea. The probability of casualty crash occurrence was calculated for a total of 35 sections of the Yeongdong Expressway (E-direction) and confirmed whether actual casualties occurred. The list of the top 10 sections for probability of casualty crash occurrence is shown in <Table 5-11>.

<Table 5-11> Top 10 sections for probability of casualty crash occurrence

No.	Section	No. Casualty	Prob.	SOR	SV	TVL	HVR	CR
1	Ansan IC - Ansan JC	0	83.72	31.31	29.39	9.85	36.00	75.20
2	Manjon JC - Wonju JC	1	83.22	86.70	124.77	8.59	37.00	50.26
3	Hobeop JC - Icheon IC	4	79.37	65.46	41.65	9.62	38.00	25.87
4	Gunja JC - Gunja TG	0	75.69	14.85	28.01	9.79	36.00	71.88
5	Myeonon IC - Pyeongchang IC	0	73.65	83.14	62.78	9.01	30.00	56.49
6	Bugok IC - N.Suwon IC	3	72.07	0.66	26.06	10.11	30.00	86.27
7	Yangji IC - Deokpyeong IC	2	70.80	57.86	20.00	9.53	26.00	74.93
8	Mumak IC - Manjon JC	7	69.89	49.33	39.28	9.06	35.00	59.74
9	Yeoju IC - Mumak IC	6	67.60	47.21	34.76	9.10	35.00	57.63
10	Saemal IC - Dunnae IC	6	66.35	35.23	38.69	8.99	30.00	80.66

Among the top 10 sections, three sections have not experienced actual casualty crashes in the last three years (2016-2018), including Ansan IC - Ansan JC, Gunja JC - Gunja TG, and Myeonon IC - Pyeongchang IC, while the other sections have actual casualty crashes. In other words, about 70% of them can be found to match the probability of casualty crash occurrence.

In addition, the list of the top five sections with a high number of casualties over the last three years (2016-2018) is shown in <Table 5-12>.

<Table 5-12> Top 5 sections for frequency of casualty crashes

Ranking	Section	No. Casualty	Notes
1	Daegwallyeong IC - Gangneung JC	9	SSES
2	Mumak IC - Manjon JC	7	-
3	E.Dunnae Hi - Myeonon IC	7	SSES
4	Yeoju IC - Mumak IC	6	-
5	Saemal IC - Dunnae IC	6	-

Of these five sections, Mumak IC - Manjon JC, Yeoju IC - Mumak IC, and Saemal IC - Dunnae IC sections can also be confirmed by the probability of casualty crash occurrence calculated based on the results of this study. The results of this study are suitable for predicting the actual crashes-prone sections. On the other hand, in the case of Saemal IC - Dunnae IC and E.Dunnae Hi - Myeonon IC sections, the probability of casualty crash occurrence was not included in the crash risk section, which was predicted to have a low probability of due to the installation of SSES at the end of 2018.

The result of comparison has shown that the crash risk sections of the development model and the actual sections of multiple crash occurrences were quite similar. Therefore, it was expected that it could be used to select candidate sites for SSES installation based on the result of this study. SSES can be installed in the sections with high probability of casualty crash based on the developed model. When selecting candidate sites for SSES installation, it may be considered to select dangerous sections based on crashes and speeding firstly, and then to install SSES in sites which traffic volume, heavy vehicle ratio, and curve ratio in the section are higher than the other sections.

6. Conclusion

6.1. Summary and Findings

The purpose of this study is to develop the prediction model of casualty crash occurrence, to quantify the effectiveness of SSES installation and to make suggestions on what needs to be considered in selecting the location for SSES installation. The main results of study conducted to achieve the objectives are as follows.

First of all, the prior study reviews for SSES installation effectiveness, installation criteria were conducted. And studies of the crash prediction model for crashes frequency and crash severity were also reviewed. In addition, the methodologies of machine learning applied in transport field were reviewed for binary classification which was used in this study. The IML which has been actively researched in recent years, was reviewed to improve predictive accuracy and interpretive performance. Through these processes, the differentiation between prior studies and this study has been clarified, and the issues that are addressed through this study has clearly been defined.

Secondly, a process of model specification was undertaken for the model development. A crash analysis before-after the installation of SSES confirmed that total crashes were reduced by about 42%, EPDO by about 71%, and casualty crash decreased by about 45%, and C-G methods were also reduced. Also, the speed analysis found that the average speed was reduced by about 7% and the proportion of exceeding speed limit decreased by about 21p% similarly by the C-G method. Next, data collection was carried out on SSES locations installed on the Korean expressways. The data of road, traffic and control conditions which were used with independent variables were collected, and basic statistics such as scatter plot, correlation analysis and box plot between variables were conducted for data's refining and filtering. In this study, variables related to speed were reflected in the development of the model through mediation effect analysis, and variables related to crash were utilized as response variables. Through analyzing the mean and standard deviation of the number of crashes, the occurrence of casualty crash was confirmed as the response variable for the model development. Therefore, the machine learning model for binary classification was applied and the IML techniques that are being actively applied in recent studies have been applied to enhance predictive accuracy and interpretability.

Thirdly, prediction model for casualty crash occurrence was developed considering the SSES installation. The developed prediction model was applied with machine learning for binary classification to predict whether or not an casualty crash occurred. IML was used to improve the prediction model's performance and description. KNN, RF, and SVM were applied to black-box models, and DT and BLR were applied to interpretable models. To improve predictive accuracy, hyper-parameter tuning went through the 10-fold cross validation process. The development result of the black-box model showed that SVM with the polynomial kernel had the best prediction accuracy of

88%. DT and BLR models were applied to the development of interpretable model by utilizing the forecast result of the black-box model. The development result of the interpretable model showed BLR's prediction accuracy of 76%. In other words, IML's black-box model was developed as SVM and the interpretable model as BLR.

Fourthly, a performance evaluation was conducted against the developed IML model compared with the typical BLR model from the perspective of the PDR framework. Comparing the accuracy of IML-SVM to BLR, the IML-SVM outperformed the BLR by about 22%. And when comparing the accuracy of IML-BLR to BLR, the IML-BLR, which was applied as a interpretable model, outperformed the BLR by about 10%. The AUC value of the IML-BLR was 0.77, which was 0.16 higher than that of the BLR, and the performance of the IML-BLR model was fair. The relevancies of BLR and IML-BLR were compared in terms of the human in the loop. The result of BLR was not appropriate in terms of significance level of SV and CR. On the other hand, it was judged that the result of IML-BLR was suitable in terms of the sign of the all independent variables and the significance level.

Fifthly, based on the IML model developed, the effects of casualty crash reduction due to SSES installation was possible to be quantified through the probability formula of casualty crash occurrence, The probability of casualty crash occurrence before installing SSES was 51% and that of casualty crash occurrence after installation was 23%. Therefore, it was possible to confirm that SSES installation reduced the probability of casualty crash occurrence by about 28%. The probability of casualty crash after the installation of SSES could be confirmed by the estimated coefficient of SSES (odds ratio) in the probability equation developed in this study. In addition, the effects of SSES installation were analyzed by separating the effects of SSES installation by direct and indirect effects through the analysis of mediation effects. The proportion of indirect effects through reducing the ratio of exceeding the speed limit was about 30% and the proportion of indirect effects through reduction of speed variance was not statistically significant at the 95% confidence level.

Finally, the probability equation of casualty crash occurrence developed in this study was applied to the sections of Yeongdong Expressway to compare the crash risk section with the actual crash data to examine the applicability of the development model. The result of comparison has shown that the crash risk sections of the development model and the actual sections of multiple crash occurrences were quite similar. Therefore, it was expected that it could be used to select candidate sites for SSES installation based on the result of this study, the probability equation of casualty crash occurrence. When selecting candidate sites for SSES installation, it may be considered to select dangerous sections based on crashes and speeding firstly, and then to install SSES in sites which traffic volume, heavy vehicle ratio, and curve ratio in the section are higher than the other sections.

6.2. Further Research

The limitations of the this study and further researches to improve this research are as follows.

First of all, a model for predicting the casualty crash developed through this study was developed with the uninterrupted traffic flow including expressways as a spatial scope. Recently, the KNPA is considering installing the SSES to reduce pedestrian casualty crashes in the urban interrupted traffic flow section. In this case, there is a limit to the application of the results developed by this study. For urban areas, there are additional considerations when installing the SSES due to delays caused by signalized intersections and detours by left and right-hand turning vehicles. Therefore, further research is needed to quantify the impact of SSES installed in urban areas and to proposed the installation criteria.

Secondly, KNN, RF, and SVM were considered as black-box model for IML development in this study. Due to the limitations of data collection for model development, deep learning algorithm such as DNN, CNN which are widely accepted in recent studies, have not been applied. For deep learning methodologies, the more data you collect, the more accurate your prediction. Therefore, it is necessary to quantify the effect of installation on SSES and to implement the study using deep learning techniques such as DNN by establishing a big-data system that systemizes the collection of relevant data for advanced research on the installation criteria.

In addition, the spatial scope of data collection in the development of the model in this study is limited to the section where SSESs are installed and its opposite direction, which may limit the prediction of the probability of casualty crash occurrence for all sections. This is why there have some mis-predicted case of the assessment results for applicability of the model development to the Yeongdong Expressway. Therefore, it is deemed necessary to further study the probability of casualty crash occurrence by expanding the spatial scope of data collection as a future research project.

Thirdly, in this study, the data samples used for model development utilized data from one year before and after SSES was installed. Therefore, further study of time-series analysis is needed for quantifying the effect of SSES installation, not only one year after installation, but also for installation effects over time considering the effect such as “the regression to the mean”.

Fourthly, although the slope ratio was not used as a significant variable for model development in this study, the difference between the upward and downward slopes may be significant in terms of the likelihood of casualty crash. Therefore, for further studies, it is necessary to develop a model by dividing the variable of slope ratio into two, the upward and the downward slope ratio in the section.

Finally, in this study, a probability equation for casualty crash occurrence was developed to quantify the effects of installation of SSES. And the direct and indirect effects of SSES were also identified through a mediation effect analysis. However, the mediation effects of SSES installation were analyzed separately by dividing the proportion of exceeding speed limit and speed variation. Therefore, a systematic analysis of SSES installation and crash reduction among variables related to speed will be needed through the multiple mediation effects analysis in future research.

References

1. Alkheder S., Taamneh M. and Taamneh S. (2016). Severity prediction of traffic accident using an artificial neural network. J. Forcasting, vol. 36, no. 1, 100-108.
2. Altman, N. S. (1992). An introduction to kernel and nearest-neighbor non-parametric regression. The American Statistician. no. 46, 175 - 185.
3. Ato, M., Vallejo, G. and Ato, E. (2014). Classical and casual inference approaches to statistical mediation analysis. Psicothema, vol. 26, no. 2, 252-259.
4. Austroads (2012). Point-to-point speed enforcement.
5. Baron, J., and Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. Journal of Personality and Social Psychology, vol. 51, 1173-1182.
6. Breiman L. (2001). Random Forests. Machine Learning, vol. 4, issue 1, 5-32.
7. Breiman, L. (2004). Consistency for a simple model of random forests. Technical Report 670, Univ. California, Berkeley, CA.
8. Carvalho, D. V., Pereira, E. M., and Cardoso, J. S. (2019). Machine learning interpretability: A survey on methods and metrics. Electronics, 8(8):832.
9. Cascetta, E., & Punzo, V. (2011). Impact on vehicle speeds and pollutant emissions of an automated section speed enforcement

system on the Naples urban motorway, In TRB 2011 Annual Meeting, vol. 17.

10. Cascetta E., Punzo, V. and Montanino, M. (2011). Empirical Analysis of Effects of Automated Section Speed Enforcement System on Traffic Flow at Freeway Bottlenecks. Transportation Research Record: J. TRB., vol. 2260, 83-93.
11. Chang L. and Wang H. (2006). Analysis of traffic injury severity: an application of non-parametric classification tree techniques. Accident Analysis & Prevention, vol. 38, issue 5, 1019-1027.
12. Christodoulou E., Ma J., Collins G. S., Steyerberg E. W., Verbakel J. Y. and Calster B. V. (2019). A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. Journal of clinical epidemiology, vol. 110, 12-22.
13. Chung J. and Sohn K. (2018). Image based learning to measure traffic density using a deep convolutional neural network. IEEE Trans. Intell. Transp. Syst. vol. 19, issue 5, 1670-1675.
14. Cox, M. G., Kisbu-Sakarya, Y., Miočević, M. and MacKinnon, D. P. (2014). Sensitivity Plots for Confounder Bias in the Single Mediator Model. Evaluation Review, vol. 37, 405-431.
15. Diao Z., Wang X., Zhang D., Liu Y., Xie K. and He S. (2019). A hybrid model for short-term traffic flow prediction for a non urban highway using artificial neural network. IEEE Trans. Intel. Transp. Syst., vol. 20, no. 3, 935-946.
16. Doshi-Velez, F. and Kim, B. (2017). Towards a rigorous science of interpretable machine learning.

17. Du M., Liu N. and Hu X. (2020). Techniques for Interpretable Machine Learning. Communications of the ACM, vol. 63, no. 1, 68-77.
18. Feng J., Wang Y., Peng J., Sun M., Zeng J. and Jiang H. (2019). Comparison between logistic regression and machine learning algorithms on survival prediction of traumatic brain injuries. Journal of Critical Care, vol. 54, 110-116.
19. Fink J., Kwigizile V. and Oh J. S. (2016). Quantifying the impact of adaptive traffic control systems on crash frequency and severity: Evidence from Oakland Country, Michigan. Journal of Safety Research vol. 57, 1-7.
20. Gianfranco F., Soddu S. and Fadda P. (2018). An accident prediction model for urban road networks. J. Transp. Saf. Secur., vol. 10, no. 4, 387-405.
21. http;//dss.Princeton.edu/training
22. https://cran.r-project.org
23. https://www.analyticsvidhya.com/blog/2019/08/decoding-black-box-step-by-step-guide-interpretable-machine-learning-models-python/
24. https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm
25. Imai, K., Keele, L., and Tingley, D. (2010). A general approach to causal mediation analysis. Psychological Methods, vol. 15, no. 4, 309-334.
26. Imai K., Keele L. and Yamamoto T. (2010). Identification, inference, and sensitivity analysis for causal mediation effects. Statistical Science vol. 25, no. 1, 51-71.
27. Imai, K., Keele, L., and Yamamoto, T. (2013). Experimental designs for identifying causal mechanisms. Journal of the Royal Statistical Society, vol. 176, Part I, 5-51.

28. Iranitalab A. and Khattak A. (2017). comparison of four statistical and machine learning methods for crash severity prediction. Accident: analysis and prevent, vol. 108, 27-36.
29. Jung Y. I., Baek T. H., Kim Y. H. and Park B. H. (2014). Traffic Accident Reduction Effects of Section Speed Enforcement System(SSES) Operation in Freeways. J. Korean Soc. Transp., vol. 32, no. 2, 119-129.
30. Kim D. Y., Lee H. W. and Hong K. S. (2019). A Study on Effectiveness Analysis and Development of an Accident Prediction Model of Point-to-Point Speed Enforcement System. Journal of the Korean Society of Safety, vol. 34, no. 5, 144-152.
31. Kim J. H., Kim K. H., Kim J. W. and Lee S. B. (2008). Estimation of Accident Effectiveness Based Upon the Location of Traffic Signal Using C-G Method. J. Korean Soc. Civ. Eng., vol. 28, no. 6, 775-789.
32. Kim J., Lee J., Park E., Kim J., Kim H., Lee J., and Jeong H. (2013). Optimal feature selection for pedestrian detection based on logistic regression analysis. Proc. IEEE Int. Conf. Syst., Man, Cyber., Manchester, U.K., 239-242.
33. Lee D. M., Kim D. H. and Song K. S. (2011). Analysis of Effects from Traffic Safety Improvement on Roadways using C-G Method. J. Korean Soc. Transp., vol. 29, no. 3, 31-40.
34. Lee H. W., Joo D. H., Hyun C. S., Jung J. H., Park B. H. and Lee C. K. (2013). A Study on the Analysis for the Effects of the Section Speed Enforcement System at the Misiryeong tunnel

section. J. Korea Inst. Intel. Transp. Syst., vol. 12, no. 3, 11-18.
35. Liu H. X., Recker W. and Chen A. (2004). Uncovering the contribution of travel time reliability to dynamic route choice using real-time loop data. Transportation Research Part A, vol. 38, 435 - 453.
36. Lynch M., White M. and Napier R. (2011). Investigation into the Use of Point-to-Point Speed Cameras, NZ Transport Agency Research Report, no. 465.
37. Ma x., Tao Z., Wang Y., Yu H.and Wang Y. (2015). Traffic speed prediction using remote microwave sensor data: Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. Transportation Research Part C: Emerging Technologies, vol. 54, 187-197.
38. Molnar C. (2018). Interpretable machine learning: A guide for making black-box models explainable.
39. Montella A., Imbriani L. L., Marzano V. and Mauriello F. (2015). Effects on speed and safety of point-to-point speed enforcement systems: Evaluation on the urban motorway A56 Tangenziale di Napoli. Accident Analysis and Prevention, vol. 75, 164-178.
40. Montella, A., Persaud, B., D'Apuzzo, M., and Imbriani, L. (2012). Safety evaluation of automated section speed enforcement system. Transportation Research Record: J. TRB, vol. 2281, 16-25.
41. Murdoch W. J., Singh C., Kumbier K., Abbasi-Asl R. and Yu B. (2018). Interpretable machine learning: definitions, methods, and applications. arXiv:1901.04592 [stat.ML], 1-11.
42. Najaf P., Duddu V. R. and Pulugurtha S. S. (2018). Predictability

and interpretability of hybrid link-level crash frequency models for urban arterials compared to cluster-based and general negative binomial regression models. International Journal of Injury Control and safety promotion, vol. 25, 3-13.

43. Norwegian Public Roads Administration (2011). Automatic Section Speed Control.
44. NSW Government (2012). Annual NSW Speed Camera Performance Review.
45. Olutayo V. A. and Eludire A. A. (2014). Traffic accident analysis using decision trees and Neural Networks. International Journal of Information Technology and Computer Science, vol. 2, 22-28.
46. Park J. J., Lee Y. M., Park J. B. and Kang J. K. (2008). The effect of Point to Point Speed Enforcement System on Traffic Flow Characteristics. J. Korean Soc. Transp., vol. 26, no. 3, 85-95.
47. Popoola O. M., Abiola O. S., Odunfa S. O. and Ismaila S. O. (2017). Comparison of raod traffic accident prediction models for two-lane highway integrating traffic and pavement condition parameters. Journal of Natural Science Engineering and Technology, vol. 16, no. 2, 1-10.
48. Rezapour M., Wulff S. S. and Ksaibati K. (2019). Examination of the severity of two-lane highway traffic barrier crashes using the mixed logit model. Journal of Safety Research, vol. 70, 223-232.
49. Rosenbaum, P. R. (2002). Observational Studies. Springer-Verlag, New York, 2nd edition.
50. Sameen M. I. and Pradhan B. (2017). Severity prediction of traffic

accidents with recurrent neural networks. Article of applied sciences, 1-17.

51. Scornet B. E., Biau G. and Vert J. (2015). Consistency of random forests. The Annals of Statistics, vol. 43, no. 4, 1716-1741.
52. Seo M. K. (2014). R for practical data analysis.
53. Shim J., Park S. H., Chung S. and Jang K. (2015). Enforcement avoidance behavior near automated speed enforcement areas in Korean expressways. Accident Analysis and Prevention, vol. 80, 57-66.
54. Hicks R. and Tingley D. (2011). Causal mediation analysis. The Stata Journal 11, no. 4, 605-619.
55. Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2013). Mediation: R package for causal mediation analysis. Journal of Statistical Software.
56. Torre L. F., Meocci M. and Nocentini A. (2019). Safety effects of automated section speed control on the Italian motorway network. Journal of Safety Research, vol. 69, 115-123.
57. Torre F. L., Meocci M., Domenichini L., Bradnzi V., Tanzi N. and Paliotto A. (2019). Development of an accident prediction model for Italian freeways. Accident analysis & prevention, vol. 124, 1-11.
58. TSO of UK (2004). Handbook of Rules and Guidance for the National Safety camera Programme for England and Wales for 2005/06.
59. TSO of UK (2007). Use of Speed and Red-Light Cameras for Traffic Enforcement: Guidance on Deployment, Visibility and Signing.
60. Wang Y. and Zhang W. (2017). Analysis of roadway and environment factors affecting traffic crash severities. Transportation Research

Procedia, vol. 25, 2119-2125.

61. Wang Y., Szeto W. Y., Han K. and Friesz T. L. (2018). Dynamic traffic assignment: A review of the methodological advances for environmentally sustainable road transportation applications. Transportation Research Part B: Methodological, vol. 111, 370-394.
62. Yin T., Zhong G., Zhang J., He S. and Ran B. (2017). A prediction model of bus arrival time at stops with multi-routes. Transportation Research Procedia, vol. 25, 4623-4636.
63. Yu B., Yand Z. Z., Chen K. and Yu B. (2010). Hybrid model for prediction of bus arrival times at next station. Journal of Advanced Transportation, vol. 44, issue 3.
64. Yun I. S. (2011). Study of the Effect of the Point-to-Point Speed Enforcement System Using a Comparison-Group Method. Journal of the Korean Society of Road Engineers, vol. 13, no. 4, 177-185.

국문 초록

Interpretable Machine Learning을 활용한 구간단속시스템 설치에 따른 인명피해사고 감소 효과 연구

서울대학교 대학원
공과대학 건설환경공학부
홍 경 식

본 연구에서는 구간단속시스템(Section Speed Enforcement System, SSES) 설치 효과를 정량화하기 위해 인명피해사고 예측모형을 개발하고, 매개효과 분석을 통해 SSES 설치에 대한 직접효과와 간접효과를 구분하여 정량화하였다. 또한, 개발한 예측모형에 대한 고속도로에서의 적용 가능성을 검토하고, SSES 설치 대상지 선정 시 고려해야할 사항을 제안하였다. 모형 개발에는 인명피해사고 발생 여부를 종속변수로 하는 이진분류형 기계학습을 활용하였으며, 기계학습 중에서는 모형의 예측 성능과 더불어 예측 결과에 대한 해석력을 중요하게 고려하는 인터프리터블 머신 러닝(Interpretable Machine Learning, IML) 방법론을 적용하였다.

IML은 블랙박스 모델과 인터프리터블 모델로 구성되며, 본 연구에서는 블랙박스 모델로 KNN, RF 및 SVM을, 인터프리터블 모델로 DT와 BLR을 검토하였다. 모형 개발 시에는 각 기법에서 튜닝이 가능한 하이퍼 파라미터에 대하여 교차검증 과정을 거쳐 최적화하였다. 블랙박스 모델은 폴리노미얼 커널

트릭을 활용한 SVM을, 인터프리터블 모델은 BLR을 적용하여 인명피해 사고 발생 확률을 예측하는 모형을 개발하였다. 개발된 IML 모델에 대해서는 PDR(Predictive accuracy, Descriptive accuracy and Relevancy) 프레임워크 관점에서 (typical) BLR 모델과 비교 평가를 진행하였다. 평가 결과 예측 정확도, 해석 정확도 및 인간의 이해관점에서의 적합성 등에서 모두 IML 모델이 우수함을 확인하였다.

또한, 본 연구에서 개발된 IML 모델 기반의 인명피해사고 발생 확률식은 SSES, SOR, SV, TVL, HVR 및 CR의 독립변수로 구성되었으며, 이 확률식을 기반으로 SSES 설치에 대한 효과를 정량화하였다. 정량화 분석 결과, SSES 설치로 인해 약 28% 정도의 인명피해사고 발생 확률이 감소함을 확인할 수 있었다. 또한, 모형 개발에 활용된 변수 중 SSES 설치로 인해 영향을 받는 변수들(SOR 및 SV)에 대한 매개효과 분석을 통해 SSES 설치로 인한 인명피해사고 감소 확률을 직접효과와 간접효과를 구분하여 제시하였다. 분석 결과, SSES와 제한속도 초과비율(SOR)의 관계에서 있어서는 약 30%가 간접효과이고, SSES와 속도분산(SV)의 관계에 있어서는 매개효과가 통계적으로 유의하지 않음을 확인할 수 있었다.

마지막으로 영동고속도로를 대상으로 인명피해사고 발생 확률식 기반의 예측 위험구간과 실제 인명사고 다발 구간에 대한 비교 분석을 통해 연구 결과의 활용 가능성을 확인하였다. 또한, SSES 설치 대상지 선정 시에는 사고 및 속도 분석을 통한 위험구간을 선별한 후 교통량(TVL)이 많은 곳, 통과차량 중 중차량 비율(HVR)이 높은 곳 및 구간 내 곡선비율(CR)이 높은 곳을 우선적으로 검토하는 것을 제안하였다.

주요어 : 구간단속시스템, 매개효과, 사고예측모형, 이진분류, 인명피해사고, 인터프리터블 머신러닝
학 번 : 2010-31011