ภาคผนวก

ภาคผนวก ก

งานประชุมวิชาการ

ATRANS

Asian Transportation Research Society

CERTIFICATE OF PARTICIPATION

This is to certify

Ms. Chatcha Kummoon

Presented a paper titled

Factors affecting pedestrians' violations of traffic signals
at mid-block Smart Crosswalk in Thailand

At the 17th ATRANS ANNUAL CONFERENCE; YOUNG RESEARCHER'S FORUM 2024
Given on the 30th of August 2024

DR. CHULA SUKMANOP
ATRANS - Chairperson

DR. TUENJAI FUKUDA
ATRANS Secretary-General,
Chair of ATRANS Conference
Organizing Committee

AYRF2024-027 (P)

ภาคผนวก ข

บทความวิชาการที่ได้รับการตีพิมพ์เผยแพร่ในระหว่างศึกษา

รายชื่อบทความวิชาการที่ได้รับการตีพิมพ์เผยแพร่ในระหว่างศึกษา

KUMMOON, Ch., PUEBOOBPAPHAN, S., & PUEBOOBPAPHAN, R. (2024) FACTORS AFFECTING PEDESTRIANS' VIOLATIONS OF TRAFFIC SIGNALS AT MID-BLOCK SMART CROSSWALK IN THAILAND. Paper presented at 17th ATRANS Annual Conference: Young Researcher's Forum (AYRF) 2024, 30 August 2024, Bangkok, Thailand. Pp.146-154

17 ATRANS ANNUAL CONFERENCE
PROCEEDINGS OF 2024
YOUNG RESEARCHER'S FORUM

Transportation for A Better Life:
"Managing Transport Decarbonization"

30 August 2024, Bangkok, Thailand

FACTORS AFFECTING PEDESTRIANS' VIOLATIONS OF TRAFFIC SIGNALS AT MID-BLOCK SMART CROSSWALK IN THAILAND

Topic number: (3) Paper Identification number: AYRF2024-027
Chatcha KUMMOON[1], Suthatip PUEBOOBPAPHAN[2], Rattaphol PUEBOOBPAPHAN[3*]

School of Transportation Engineering, Institute of Engineering
Suranaree University of Technology
Telephone: 061-0281370[1], 044-224645[2], 044-224593[3*]
E-mail: M6402662@g.sut.ac.th[1], suthatip@sut.ac.th[2], rattaphol@sut.ac.th[3*]

Abstract

This study examines the factors that influence instances of pedestrians disobeying traffic signals at mid-block crossings that are equipped with an intelligent crossing device, known as the Smart Crosswalk in Thailand. Once fully comprehended, it is postulated that recommendations or interventions to address pedestrian signal violations can be formulated to mitigate pedestrian accidents at crossings nationwide. The analysis examines several factors, including highway-related geometry, traffic volume, speed, and system operational parameters. The dependent variable is the binary event of whether the pedestrian signal is disobeyed or not during each cycle. Binary Logistic Regression is utilized to create models for this purpose. As the number of signal violation events is far lower than that of not being violated, the data in this study is considered imbalanced. Therefore, some scaling techniques to deal with imbalanced data are investigated and the downsampling technique is adopted for the final model development. This approach resulted in the highest Area Under the Curve (AUC) value of 71.70%. The specificity, sensitivity, and accuracy of the model were as follows: 67.56%, 68.29%, and 67.61%, respectively. The results show that while pedestrians are more likely to violate pedestrian signals as waiting time grows, the opposite is true when crossing length, traffic volume, and speed increase. Field observation reveals that violations are more likely at night when traffic volumes are low and signal waiting times are long.

Keywords: Pedestrians, Pedestrian safety, Signal violation, Smart crosswalk, Binary Logistic Regression, Imbalanced data

1. Introduction

Road accidents are one of the main causes of death in the world. According to a report by the World Health Organization on Traffic Injuries every year, road accidents result in more than 1.3 million deaths worldwide per year, and more than half of all deaths and injuries involve vulnerable road users such as pedestrians, cyclists, motorcyclists, and passengers[1]. According to the WHO's 2018 Global Situation Report on Road Safety, Thailand has the highest death rate from road accidents in Southeast Asia and the 9th highest in the world, with 32.7 deaths per 100,000 people. Motorcycles had the highest death rate at 74%, followed by pedestrians at 8%[2]. Out of the total number of road accidents, an average of six to eight pedestrian traffic accidents occur each day[3]. In Thailand, road accidents kill two pedestrians on average each day. Although this is a small number compared to other types of road users, what is worrisome is the increasing trend in the severity of pedestrian accidents. Analysis of the severity index based on highway accident statistics from the Department of Highways revealed that pedestrian accidents were approximately four times more severe than general accidents on highways. Consequently, once a pedestrian accident occurs, it leads to an exceptionally high fatality rate[4]. The primary cause of pedestrian fatality or injury, as determined by global studies, is being hit by a vehicle while pedestrians are crossing the roadway[5].

The Department of Highway (DOH) has implemented the Smart Crosswalk system to enhance pedestrian safety at mid-block crossings. This system incorporates pedestrian and vehicle detection devices to analyze the requirements for using the crossing and regulate the signal lights, thereby ensuring optimal safety and efficiency. Despite the installation of the intelligent crossing technology, pedestrian violations are still evident.

Previous research has primarily employed binary logistic regression analysis to investigate factors affecting pedestrian traffic signal violations. For example, research in Kolkata, India focused on various elements influencing pedestrians' red-light violations at three intersection crosswalks in different cities. The findings revealed that non-social factors such as crossing speed, waiting time for crossing signals, traffic volume, and signal cycle length significantly impact signal violation behavior. Social factors, including the number of people waiting at intersections and observing others successfully violating the red light, also positively influence this behavior[6]. In Hong Kong, the impact of personal factors and environmental factors on the individual decision of a red-light running violation was examined at the signalized crossings in the urban area. The results indicated that pedestrian gender, age, number of lanes, presence of a companion, number of pedestrians around, presence of other violators in the same cycle, time to green, red time, traffic volume, and percentage of heavy vehicles all significantly impact the propensity of red-light running violation of pedestrians[7]. Similarly, research in Mashhad, Iran, evaluated the impact of external factors on pedestrian violations at traffic signal intersections. The findings indicated that traffic volume, the number of violators, crosswalk length, red light duration (waiting time), and physical movement problems influence pedestrians' decisions to violate traffic signals[8]. Prior research reviews showed that most studies, both domestically and internationally, took place at crosswalks with fixed-time signals. Currently, there is a lack of research to investigate and comprehend the factors that influence pedestrians' violation behavior at crosswalks with smart crosswalk systems.

Our preliminary analysis of pedestrian violation at smart crosswalks indicates that the data in this study is imbalanced, i.e. the number of events that pedestrian violations are observed during any signal cycle is far less than the events of no violation. If the model is developed without addressing unbalanced property of the data, the model's predicted outputs might be biased. Noppamas[9] conducted research on comparative analysis to evaluate the effectiveness of Random OverSampling Examples (ROSE) and Synthetic Minority Oversampling Technique (SMOTE) in addressing data imbalances. The results showed that SMOTE outperformed both the initial data set and ROSE. Kesornsit[10] compares the effectiveness of a variety of methods for addressing the issue of imbalanced data in diabetic patients. Addressing the data imbalance can improve classification efficiency, making it more efficient than the initial data and enabling the model to classify both large and small groups of data equally or nearly equally. Previous research conducted in Thailand predominantly employed imbalance adjustments with medical data, financial data, demographic data, educational data, etc. In the context of pedestrian violation data at crosswalks with traffic signals or with smart crosswalk systems, no prior research has adjusted for imbalanced data.

Based on the aforementioned research gaps, this study investigates the factors influencing pedestrians' violations of traffic signals at mid-block smart crosswalks in Thailand by taking imbalanced property of the data into consideration. The analysis of the smart crosswalk system relies on secondary data, including traffic volume, speed, operational parameters, and road characteristics. However, the secondary data from the smart crosswalk system lacks information about the individual characteristics of pedestrians, such as gender, age, and occupation. Therefore, this study cannot consider factors related to individual characteristics. Additionally, the pedestrian violation data recorded in the system is the total number of pedestrians that violated signal for each cycle. Therefore, the dependent variable employed in this study is binary, indicating whether pedestrian signal is violated for each signal cycle. Binary Logistic Regression is therefore chosen to perform the analysis. The primary objective is to generate recommendations or measures to address signal violations and reduce pedestrian accidents.

2. Data collection

Currently, there are 11 Smart crosswalks in Thailand (Fig. 1). For this study, we selected six crossings: 1) Highway No. 3242, km 18+110, Bangkok; 2) Highway No. 306, km 4+970, Nonthaburi Province; 3) Highway No. 3316, km 2+225, Nakhon Pathom Province; 4) Highway No. 3215, km 11+020, Nonthaburi Province; 5) Highway No. 3242, km 2+220, Samut Sakhon Province; and 6) Highway No. 314, km 8+410, Chachoengsao Province. We did not consolidate the data from the other five crossings for analysis, as some of the systems were temporarily down and others had only recently been implemented with a smart crosswalk.

This study utilized data on physical road characteristics, such as the median, median width, lanes, crosswalk width, crosswalk length, yellow frame, and yellow frame width. Secondary data from smart crosswalk systems, including speed and traffic volume, as well as specifications related to traffic signal systems at crossings, were also used. The study exclusively focused on factors anticipated to have a direct correlation with pedestrians violating traffic signals, considering the system’s definition of violating pedestrians, the correlation coefficient, and previous literature. Variables not listed above, such as characteristics of individual pedestrian and pedestrian behavior, were not considered as they are not available from the system database.

We collected data from the smart crosswalk system from January 2022 to August 2022.

Fig. 1 Smart crosswalks in Thailand

2.1 Physical Characteristics of Smart Crosswalks

We conducted a survey of the crossing's physical characteristics, encompassing three main aspects: 1) the road's physical attributes, including the presence or absence and width of the median and the number of traffic lanes; 2) the crossing's physical features, such as its length and width; and 3) the safety equipment at the crossing, including the presence or absence and width of the yellow netting frames (Table 1).

2.2 Smart Crosswalk System database

The Smart Crosswalk System records a number of databases, including traffic, speed, pedestrian, and transaction information. For this study, we focus on the traffic and transaction databases. The traffic database provides data on traffic volume and average speed, organized into 5-minute intervals. The transaction database contains detailed information on pedestrian signals, including waiting time for crossing signals, the duration of the walk interval, the length of the flashing don't-walk signal, and any extensions of this signal. Additionally, it records the total number of pedestrians crossing and the number of pedestrians who violate the signal during each signal cycle.

The dependent variable in this study is the event of pedestrian violations occurring during signal cycles, categorized based on whether or not pedestrian violation is observed during that cycle. To mitigate issues related to multicollinearity and based on literature findings, only selected factors are considered as independent variables. Table 2 summarizes the variables included in this study.

Table 1 Physical Characteristics of Smart Crosswalks

Smart crosswalk	Lane/ direction	Crosswalk length (m)	Crosswalk width (m)	Yellow frames width (m)	Median width (m)
1) Highway No. 3242 km 18+110	3	25	8	0	1.3
2) Highway No. 306 km 4+970	2	7.8	8	0	0.8
3) Highway No. 3316 km 2+225	2	14	8	7	0
4) Highway No. 3215 km 11+020	3	23	8	7	4.3
5) Highway No. 3242 km 2+220	3	30	8	12	10
6) Highway No. 314 km 8+410	3	35	16	17.7	5

Source: Bureau of Highway Safety (2022) As Built Drawing for Implementing the Smart Crosswalk, Department of Highways, Thailand.

Table 2 Description of variables in this study

Dependent Variable	Explanation
Event of pedestrian violations that occur during signal cycles	Dichotomous Variable 0 = none of the pedestrians violated a traffic signal (Not Violated). 1 = at least one pedestrian violated a traffic signal (Violated).
Independent Variable	**Explanation**
Waiting time	The waiting time is calculated from the moment pedestrians press the button to request pedestrian signals until they receive the signal to cross (unit: seconds).
Crosswalk length	(unit : meters)
Vehicle speed	We calculate the vehicle average speed every 5 minutes, which corresponds to the signal cycle (unit : km/hr).
Traffic volume	We calculate the traffic volume every 5 minutes, which corresponds to the signal cycle (unit : vehicle/hr/lane).

3. Method

RStudio was employed to develop models using Binary Logistic Regression, and to process, analyze, and forecast data. We analyzed a total of 116,485 signal cycles, dividing them into two groups: 93,188 cycles (80%) for training data and 23,297 cycles (20%) for test data.

3.1 Binary Logistic Regression Analysis

The binary logistic regression is used for regression analysis where the dependent variable is dichotomous, taking on values of 0 or 1. This model investigates which independent variables lead to pedestrian violations during signal cycles. The analysis uses the following function to handle situations with multiple predictor variables:

$$P_y = \frac{e^{b_0+b_1x_1+\cdots+b_px_p}}{1+e^{b_0+b_1x_1+\cdots+b_px_p}} \quad (1)$$

Where: P_y = the likelihood of the event of interest. and Q_y = the likelihood of the event of interest not occurring, as follows:

$$Q_y = 1 - P_y \quad (2)$$

Given the non-linear relationship between predictor and dependent variables in logistic regression, we adjust the relationship to be linear in the form of odds ratios. The odds ratio (OR), which is the expected change in probabilities for a unit increase in the predictor, is expressed as Exp(B)[11].

$$Odds\ Ratio\ (OR) = \frac{P_y}{Q_y} \quad (3)$$

3.2 Measures of accuracy

The performance of the model can be verified using the following statistical values: True Positive (TP), False Negative (FN), False Positive (FP), and True Negative (TN), as shown in Table 3. These values are used to create metrics to evaluate the model performance using equations (4) – (6)[12]. Additionally, the Area Under the Curve (AUC) can be used as another model's performance indicator. A model with an AUC value closer to 1 indicates better predictive performance[13].

Table 3 Confusion Matrix

	Predicted Positive	Predicted Negative
Actual positive	True Positive (TP)	False Negative (FN)
Actual negative	False Positive (FP)	True Negative (TN)

Source: [12]

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \quad (4)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (5)$$

$$Specificity = \frac{TN}{FP + TN} \quad (6)$$

3.3 Imbalanced data

Imbalanced data poses a problem when the proportions of data are unequal, leading the training model to be dominated by the majority class. Consequently, the model tends to predict the majority class more accurately than the minority class, resulting in a high specificity rate but a low sensitivity rate in binary classification. Data scaling is one technique to mitigate this issue. Common scaling methods used in practice include the following[14]:

Upsampling: This method increases the size of the minority class by sampling with replacement, making the majority and minority classes equivalent in size.

Downsampling: This method reduces the size of the majority class by sampling with replacement, making the majority and minority classes equivalent in size.

Hybrid methods: Techniques such as ROSE and SMOTE downsample the majority class while adding synthesized information to the minority class from the original dataset.

These scaling methods should be applied exclusively to the training data and not to the test data.

4. Results

This study began with an analysis using descriptive statistics to examine the behavior of drivers and pedestrians, as well as the functioning of the smart crosswalk system. The results are presented in Table 4.

Table 4 indicates that the percentage of violated signal cycle was only 6.86 percent, which is relatively small, and the dataset could be defined as imbalanced. Despite the low occurrence, such violations pose a significant risk of causing serious accidents for both drivers and pedestrians. The average speed was found to be 42.57 km/hr, with the 85th percentile speed at 59.14 km/hr, which is higher than the recommended safe speed for pedestrians (30–40 km/hr)[15].

The results of the binary logistic regression analysis based on the original train dataset without addressing imbalanced feature are presented in Table 5.

Table 4 Descriptive statistics of the smart crosswalk data

Dependent variable				Frequency		Percentage	
1. Events of pedestrian violations that occur during signal cycles (n = 116,485 signal cycles)							
None of the pedestrians violated a signal cycle (Not Violated).				108,490		93.14 %	
At least one pedestrian violated a signal cycle (Violated).				7,995		6.86 %	
Independent variable	**Min**	**P15**	**Mean**	**Median**	**P85**	**Max**	**SD**
2. Waiting time (second)	8	10	22.75	20	35	65	11.54
3. Crosswalk length (meter)	13.8	13.8	22.94	23.2	35	35	7.74
4. Vehicle speed (km/hr)	5.25	24.74	42.57	43.29	59.14	115.13	16.45
5. Traffic volume (veh/hr/lane)	2	136	247	234	350	1,650	122

Table 5 Results of binary logistic regression based on original train dataset

Independent variable	Coefficient (B)	Standard Error	Exp (B)	P-value
Constant	- 1.042 ***	0.054	0.353	0.000
Waiting time (second)	0.041 ***	0.001	1.042	0.000
Crosswalk length (meter)	- 0.075 ***	0.002	0.927	0.000
Vehicle speed (km/hr)	- 0.018 ***	0.001	0.983	0.000
Traffic volume (veh/hr/lane)	- 0.001 ***	0.000	0.999	0.000

Note*** Statistical significance level at 0.001, $R^2 = 0.0859$, Specificity = 100% , Sensitivity = 0%, Accuracy = 93.14% , AUC = 71.46%

Table 5 reveals that all independent variables have a statistically significant influence on pedestrians' violations of traffic signals at the 0.001 level. However, when using the test dataset to make predictions, the model failed to accurately predict pedestrian violations of traffic signals (as sensitivity is 0.0%). This aligns with the findings of previous studies[14] on imbalanced datasets. To address this issue and improve prediction accuracy, several methods were employed, including Upsampling, Downsampling, ROSE, and SMOTE. These methods were used to adjust the number of signal cycles with pedestrian violations to be similar to the number of signal cycles without pedestrian violations.

After applying these data scaling techniques, new models were developed. The results indicated that each independent variable continued to have a statistically significant effect on pedestrian violations of traffic signals at the 0.001 level. Furthermore, the relationship between the independent and dependent variables remained consistent with the initial analysis. The prediction performance based on the binary logistic regression model from the new data scaling experiments are presented in Table 6.

Table 6 presents performance of the prediction results for pedestrian violations of traffic signals, including specificity, sensitivity, overall accuracy, and area under the curve (AUC). The application of data scaling principles across four methods resulted in improved sensitivity, which is crucial in this context, compared to the model's performance without data scaling. Based on these results, we decided to use the downsampling method to adjust the proportion of signal cycles. This method achieved the highest AUC value of 71.70%. The corresponding specificity, sensitivity, and accuracy were 67.56%, 68.29%, and 67.61%, respectively. Downsampling is particularly suitable for large datasets[16]. The results of binary logistic regression with the downsampling method, are presented in Table 7 and illustrated in Fig. 2.

Table 6 Prediction performance of scaling techniques based on test dataset

Scaling techniques	Specificity	Sensitivity	Accuracy	AUC
Upsampling	67.65%	68.48%	67.71%	71.64%
Downsampling	67.56%	68.29%	67.61%	71.70%
ROSE	67.60%	68.36%	67.65%	71.65%
SMOTE	75.06%	59.29%	73.98%	71.62%

Table 7 Results of binary logistic regression based on downsampling method

Independent variable	Coefficient (B)	Standard Error	Exp (B)	P-value
Constant	1.307 ***	0.076	3.695	0.000
Waiting time (second)	0.050 ***	0.002	1.052	0.000
Crosswalk length (meter)	- 0.079 ***	0.003	0.924	0.000
Vehicle speed (km/hr)	- 0.014 ***	0.001	0.986	0.000
Traffic volume (veh/hr/lane)	- 0.001 ***	0.000	0.999	0.000

Note*** Statistical significance level at 0.001, R^2 = 0.1232, Specificity = 67.56% , Sensitivity = 68.29% , Accuracy = 67.61% , AUC = 71.70%

Table 7 presents a model of pedestrian traffic signal violations developed using the downsampling method. This model indicates that each independent variable has a statistically significant effect on pedestrian violations at the 0.001 level. The model accounts for 12.32% of the variance in pedestrian violations. The relationships between independent and dependent variables in this model are similar to those observed without data scaling, but now with higher predictive capability. Performance of the original model and the downsampling model can be compared using confusion matrix as shown in Table 8. The downsampling model is superior to the original model in predicting true positive (TP) but at the cost of increasing false positive (FP). Explanation of the results for each variable based on the downsampling method are as follows:

Waiting time (B = 0.050, Exp(B) = 1.052: For each additional second of waiting time, while holding other factors constant, the likelihood of a pedestrian signal cycle being violated increases by 5.2%.

Crosswalk length (B = -0.079, Exp(B) = 0.924): For each additional meter of crosswalk length, while holding other factors constant, the likelihood of a pedestrian signal cycle being violated decreases by 7.6%.

Average Vehicle Speed (B = -0.014, Exp(B) = 0.986): For each additional kilometer per hour increase in average vehicle speed; while holding other factors constant, the likelihood of a pedestrian signal cycle being violated decreases by 1.4%.

Traffic Volume (B = -0.001, Exp(B) = 0.999): For each additional vehicle per hour per lane, while holding other factors constant, the likelihood of a pedestrian signal cycle being violated decreases by 0.1%.

Table 8 Performance of the original model and the downsampling model with test dataset

	Original model		Downsampling model	
	Predicted Positive	Predicted Negative	Predicted Positive	Predicted Negative
Actual positive (1,599 cycles)	0 (TP)	1,599 (FN)	1,092 (TP)	507 (FN)
Actual negative (21,698 cycles)	0 (FP)	21,698 (TN)	7,040 (FP)	14,658 (TN)

Note*** Positive Class = 1 (at least one pedestrian violated a traffic signal (Violated).)

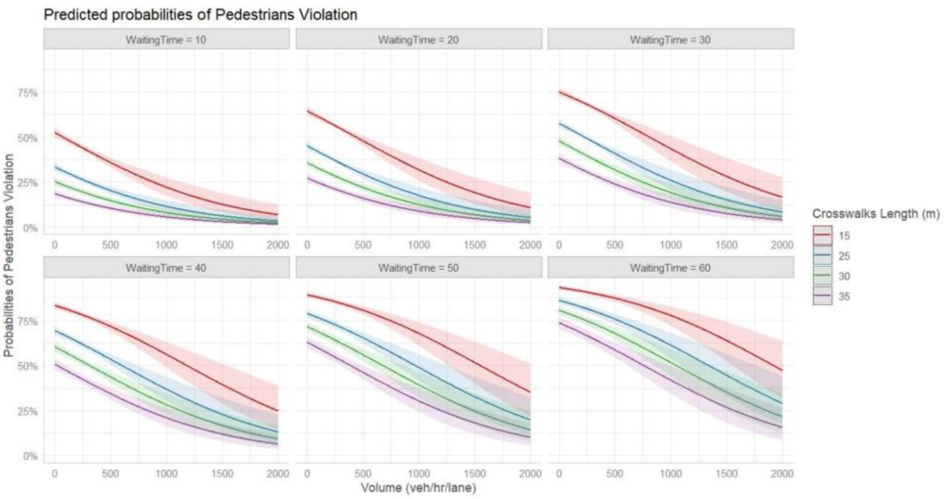

Fig. 2 Predicted probabilities of pedestrian violation

Fig. 2 presents six subgraphs, each representing different waiting times for the signal and illustrating the relationship between crosswalk length, traffic volume, and pedestrian traffic signal violations. The crosswalk lengths are categorized into 15, 25, 30, and 35 meters. It shows that violations are more likely when shorter crosswalk length, longer waiting time, and lower traffic flow. This finding aligns with field observation, which indicated that signal violations are more common at night when traffic flow is low, and pedestrians experience prolonged waiting times after pressing the button. To reduce violation rates, it is recommended to adjust the waiting times to better align with current traffic conditions in the smart crosswalk system.

5. Conclusion

This study investigates the factors influencing pedestrian violations of traffic signals at Smart Crosswalks in Thailand, using data collected from January 2022 to August 2022. The analysis employed binary logistic regression to assess the impact of four independent variables on pedestrian signal violations. Results revealed that each independent variable had a statistically significant effect on violations at the 0.001 level. However, when applying the model to the test set, it failed to effectively predict pedestrian violations. To address this limitation, the number of signal cycles without pedestrian violations was adjusted to match those with violations, enhancing the model's predictive capacity. The downsampling method, which achieved the highest Area Under the Curve (AUC) value of 71.70%, demonstrated improved performance compared to other techniques, with specificity, sensitivity, and accuracy values of 67.56%, 68.29%, and 67.61%, respectively. Despite these improvements, the model's overall predictive performance remained suboptimal. The analysis revealed a positive correlation between pedestrian violations and waiting time, and negative correlations with crossing length, traffic speed, and traffic volume, which aligns with field observations. To reduce violation rates, it is suggested that system waiting times be adjusted to better reflect traffic conditions. Despite the improved performance obtained from the data scaling methods, the specificity, sensitivity, accuracy, and AUC values remained relatively low[17]. Although binary logistic regression effectively identifies relationships between variables, it may be inadequate for precise forecasting. Future research should explore alternative machine learning techniques, such as Decision Trees, Random Forests, or Gradient Boosting, together with data scaling methods to enhance predictive performance and validate these findings.

References

[1] World Health Organization (2022) World Health Statistics 2022, Retrieved Jun 15, 2024, from: https://www.who.int/news/item/20-05-2022-world-health-statistics-2022.

[2] World Health Organization (2022) GLOBAL STATUS REPORT ON ROAD SAFETY 2018., Retrieved Jun 15, 2024, from: https://www.who.int/publications/i/item/9789241565684.

[3] National Statistical Offie (2021) Statistics on road traffic accident cases Classified by type of road user, damage, suspects, by region 2012 - 2021., Retrieved Jun 16, 2024, from: https://www.nso.go.th

[4] Thai Roads Foundation (2019) Key factor on road safety situations in thailand 2016-2017., Retrieved Jun 16, 2024, from: http://trso.thairoads.org/statistic/national

[5] Thai Health Promotion Foundation (2019) Opening the perspective to reflect the reality of vulnerable people on the road (VRUs) ., Retrieved Jun 16, 2024, from: https://data.go.th/th/dataset/rtddi

[6] Raoniar, R., & Maurya, A. K. (2022) Pedestrian red-light violation at signalised intersection crosswalks: Influence of social and non-social factors., Safety science, 147, 105583.

[7] Zhu, D., Sze, N. N., & Bai, L. (2021) Roles of personal and environmental factors in the red light running propensity of pedestrian: Case study at the urban crosswalks., Transportation research part F: traffic psychology and behaviour, 76, 47-58.

[8] Afshari, A., Ayati, E., & Barakchi, M. (2021) Evaluating the effects of external factors on pedestrian violations at signalized intersections (a case study of Mashhad, Iran)., IATSS research, 45(2), 234-240.

[9] Noppamas, D. (2018) Comparison of Imbalanced Data Problem Solvingfor Income Classification of Type I Pharmacies Entrepreneur., In The 9 th STOU National Research Conference (pp. 1578-1586).

[10] Kesornsit, W., Lorchirachoonkul, V., & Jitthavech, J. (2018) Imbalanced data problem solving in classification of diabetes patients., Khon Kaen University Research Journal, 18(3), 11-21.

[11] Kaiyawan, Y. (2012). Principle and using logistic regression analysis for research. RMUTSV Research Journal, 4(1), 1-12.

[12] lamprakhon, B. (2017) A proposed method to compare areas under the ROC cuves for a single

dataset: A case of credit scoring model., Chulalongkorn University Theses and Dissertations (Chula ETD)., from: https://digital.car.chula.ac.th/chulaetd/2020

[13] Šimundić, A. M. (2009). Measures of diagnostic accuracy: basic definitions. ejifcc, 19(4), 203.

[14] Modeling with R. (2019) Methods for dealing with imbalanced data., Retrieved Jun 15, 2024, from: https://www.r-bloggers.com/2019/04/methods-for-dealing-with-imbalanced-data/.

[15] World Health Organization. (2023) Pedestrian safety: a road safety manual for decision-makers and practitioners., Retrieved Jun 30, 2024, from: https://www.who.int/publications/i/item/9789240072497

[16] Minh, H. (2018) How to Handle Imbalanced Data in Classification Problems., Retrieved Jun 15, 2024, from: https://medium.com/@nminh.hoang1023/handling-imbalanced-data-in-classification-problems-7de598c1059f.

[17] Metz, C. E. (1978). Basic principles of ROC analysis., In Seminars in nuclear medicine (Vol. 8, No. 4, pp. 283-298). WB Saunders.