International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.

http://ijpsat.ijsht-journals.org
Vol. 29 No. 2 November 2021, pp.499-511

Estimating The Causes of Motorcycle Accidents Using Logistic Regression

N. A.M.R.Senaviratna[1]* and T.M.J.A.Cooray[2]
[1]Department of Mathematics
The Open University of Sri Lanka
Sri Lanka

[2]Retired Lecturer
Sri Lanka

Abstract— Ever increasing road accidents and traffic flow is a heavy burden to a developing country like Sri Lanka. Among road accidents, motorcycle accidents involved a high rate of accidents. Therefore, it is essential to find solutions and reduce motorcycle accident deaths and injuries. The objective of this study is to estimate the causes of motorcycle accidents in Sri Lanka. Secondary data used in this study from 2014 to 2016 were acquired from the Traffic Police headquarters, Colombo in Sri Lanka. A total number of 32926 motor cycle accidents were included in the analysis. Two third of data (21950) was used to develop the model, and the remaining 1/3 of data (10976) was used to validate the model. Factors considered in the study were gender of motorcyclist, validity of license, accident cause, alcohol test, time of accident, weekday/weekend, road surface, weather condition, light condition, location and age of motorcyclist. Severity of accidents was categorized as grievous and non-grievous accidents. Chi-square test of independence has detected that alcohol test is not significantly associated with the severity of accidents. Binary logistic regression is applied to model the severity of road accidents due to the dichotomous nature of the dependent variable. The area under the receiver operating characteristic (ROC) curve was 0.587 which means the fitted model classifies the group significantly better than by chance. The fitted model is correctly predicted 79.6% of the validation data which is greater than the predictive power of the baseline model 56.6%. Results revealed that location type, time, age of driver, accident cause and gender have a significant effect on the severity of accidents. Moreover, Aggressive or negligent driving, driving on straight road, drive by male motorcyclists, driving in daytime have a high chance to be a grievous accident. These findings can aid in modifying laws and establishing preventive approaches in Sri Lanka.

Keywords—motorcycle accidents; logistic regression; severity of accident; grievous accidents; non-grievous accidents.

I. INTRODUCTION

Motorcycles are the vehicles which offer a greater sense of freedom and openness than any other vehicle. Being more fuel efficient than other vehicles, motorcycles are more preferred by people as a cost effective mode of transportation in developing countries like Sri Lanka. Motorcycle accidents are most commonly serious and hazardous that can not only cause serious damage in injuries but also can take away the life. In the year 2016, 10754 motorcycle accidents were reported where 11% of them are fatal contributing to 1178 deaths. Motorcycles can be anticipated throughout the country which will result in more road casualties and tremendous economic losses, especially the extra health care costs for the accident victims and therefore, remains a challenging issue to all concerned parties to address this significant social problem and concurrently, to implement all the necessary measures promptly to fight this long and seemingly endless battle [17]. Therefore, it is important to find the risk factors associated with motorcycle accidents. The objective of this study is to estimate the causes of motorcycle accidents in Sri Lanka.

II. Materials And Methods

2.1. Binary logistic regression

Binary logistic regression model estimates the probability of occurrence of an event by fitting data to a logistic curve. The dependent variable is the population proportion or probability that the resulting outcome is equal to 1. Parameters obtained for the explanatory variables can be used to estimate odds ratios for each of the explanatory variables in the model.

The specific form of the logistic regression model is:

$$\pi(x) = \frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots\beta_p x_p}}{1+e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots\beta_p x_p}} \qquad (1)$$

where π is the probability of the outcome of interest or event, β_0 is the intercept, $\beta_1, \ldots, \beta_p$ are regression coefficients, $x_1, x_2, \ldots, x_p$ are independent variables.

The transformation of the conditional mean π(x) logistic function is known as the logit transformation:

$$ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots. + \beta_p x_p \qquad (2)$$

The importance of the logit transformation is that it is linear in its parameters and may range from. $-\infty$ to $+\infty$

There are four assumptions under binary logistic regression. These assumptions are required to satisfy to give a valid result.

- Linearity: The assumption of linearity in logistic regression is that any explanatory variables have a linear relationship with the logit of the response variable. If the relationship between the log odds of the response occurring and each of the explanatory variables is not linear then the model will not be accurate.
- Independent errors: the assumption of independent errors states that errors should not be correlated for two observations.
- Multicollinearity: The assumption requires that explanatory variables should not be highly correlated with each other.
- There should be no outliers, high leverage values or highly influential points.

2.2. Source of data

Secondary data used in this study from 2014 to 2016 were acquired from the Traffic Police headquarters, Colombo in Sri Lanka. Main dataset (32926) was divided into two portions; 2/3 of data (21950) was used to develop the model, and the remaining 1/3 of data (10976) was used to validate the model. Factors considered in the study were the gender of a motorcyclist, the validity of the license, accident cause, alcohol test, time of the accident, weekday/weekend, road surface, weather condition, light condition, location and the age of motorcyclist. Response variable is accident severity which consists of two levels namely grievous and non-grievous. Accidents result in death or critically injured are named as grievous accidents and accidents result in non-critically injured or damage only accidents are named as non-grievous accidents.

2.3. Research Methodology

Data analyses are arrayed mainly under preliminary and fundamental analyses. Under preliminary analysis, descriptive statistics was performed to get a general understanding of the whole data set. In fundamental analysis, Pearson Chi-Square test is performed to check the association between each contributory factor and the severity of accidents. Then multicollinearity is checked and highly correlated variables were removed from the analysis. Finally, due to the dichotomous nature of the dependent variable, binary logistic regression analysis is carried out as advanced analysis to investigate the effect of the variables. These statistical data analyses are conducted by using EVIEWS and SPSS software.

III. Results And Discussion

3.1. Severity of Motorcycle Accidents

Following Figure 1 presents grievous and non-grievous accidents occurred during the period 2014-2016.

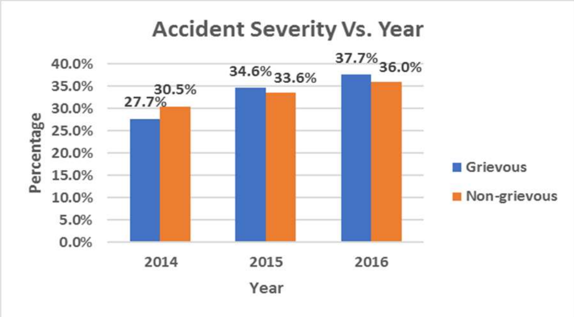

Fig 1: Percentage of accident severity vs. year

Figure 1 illustrates that percentage of grievous accidents is increased from 28 to 38 from 2014 to 2016. Non-grievous accident percentage is increased from 30 to 36 from 2014 to 2016. Percent of grievous accident is increased by 10% more in 2016 than in 2014.

3.2. Descriptive Statistics

Table 1 shows the distribution of factors considered in this study. It can be seen that majority of accidents were occurred in roads with dry surface, clear weather under daylight. The cause of most of accidents was aggressive or negligent driving.

TABLE 1: DESCRIPTION OF FACTORS

Factor	Levels	Non-grievous (%)	Grievous (%)	Total (%)	Abbreviation
Gender	Male	13691(42.6%)	18450(57.4%)	32141(100%)	M
	Female	282(35.9%)	503(64.1%)	785(100%)	F
Validity of License	With valid license	7586(38.2%)	12263(61.8%)	19849(100%)	WL
	Without valid license	6387(48.8%)	6690(51.2%)	13077(100%)	WOL
Accident Cause	Speeding	2551(48.9%)	2663(51.1%)	5214(100%)	Cause1
	Aggressive/negligent driving	9741(41.1%)	13938(58.9%)	23679(100%)	Cause2
	Influenced by alcohol/drugs	750(41.7%)	1047(58.3%)	1797(100%)	Cause3
	Others	931(41.6%)	1305(58.4%)	2236(100%)	Others
Alcohol Test	No alcohol/below legal limit	13083(42.5%)	17731(57.5%)	30814(100%)	BL
	Over legal limit	890(42.1%)	1222(57.9%)	2112(100%)	OL
Time	Day time	8294(39.5%)	12705(60.5%)	20999(100%)	DT
	Night time	5679(47.6%)	6248(52.4%)	11927(100%)	NT
Weekday/Weekend	Weekday	9541(41.9%)	13212(58.1%)	22753(100%)	WD
	Weekend	4432(43.6%)	5741(56.4%)	10173(100%)	WE

Road surface	Dry	13287(42.3%)	18135(57.7%)	31422(100%)	D
	Wet	642(45.8%)	759(54.2%)	1401(100%)	W
Weather Condition	Clear	13302(42.3%)	18157(57.7%)	31459(100%)	CL
	Rainy	393(44.6%)	488(55.4%)	881(100%)	RA
Light condition	Daylight	8102(39.4%)	12440(60.6%)	20542(100%)	DL
	Night, no street lighting	4759(48.6%)	5024(51.3%)	9783(100%)	NSL
	Others	1112(42.7%)	1489(57.2%)	2601(100%)	DD
Location	Bend/Junction	3017(38.1%)	4894(61.9%)	7911(100%)	BJ
	Road	10956(43.8%)	14059(56.2%)	25015(100%)	RD

3.3. Distribution of Age of Driver

Driver's age is the only one continuous variable in this data set. The distribution and descriptive statistics of age of drivers at fault is shown in the following Figure 2 and Table 2 respectively.

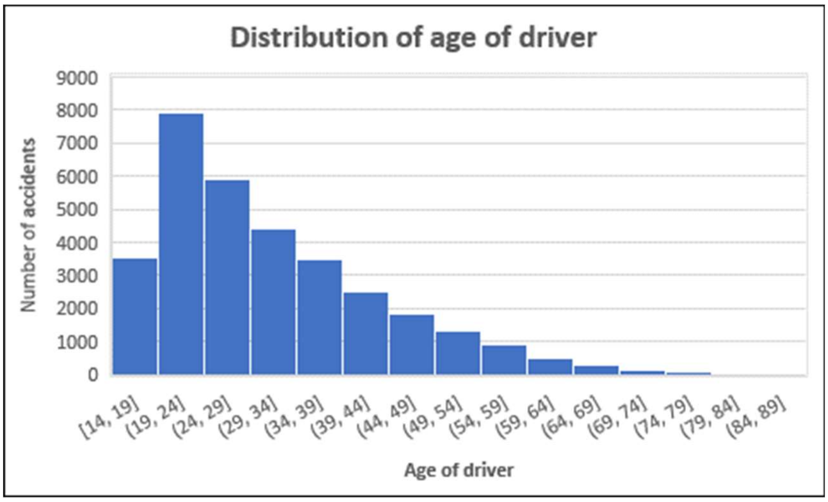

Fig 1: Distribution of age of driver

TABLE 2: DESCRIPTIVE STATISTICS OF AGE OF DRIVER

	Mean	Std. Deviation	Minimum	Maximum	Q1	Q3
Age of Driver	31.7	11.95	14	88	22	38

Table 2 shows that average age of driver at fault for accident is 31.7 years. Furthermore, even though valid age for issuing a driving license is 18 years, above Table indicates that minimum age of driver at fault for accident is 14 years. To get an idea of which age range of drivers are more responsible for accidents, histogram is drawn as above.

According to the Figure 2, the drivers in between the age group 19-24 were more responsible for the highest number of accidents which is approximately 8000 accidents. Furthermore, it indicates that younger drivers whose age less than 30 years were accounted for 60% of total accidents. it should be noted that younger drivers were more responsible for the highest number of motorcycle accidents.

3.4. Results of Pearson Chi-Square Test

Then Pearson Chi-square test is performed to test the significant relationship between the independent variables and the dependent variable. According to this analysis, following table describes the association between each factor and the accident severity.

TABLE 3: RESULTS OF CHI-SQUARE TEST

Variables	**χ^2 value**	**P value**
Gender of Driver	$\chi^2(1) = 13.969$	0.000
Validity of License	$\chi^2(1) = 364.172$	0.000
Alcohol Test	**$\chi^2(1) = 0.082$**	**0.775**
Accident Cause	$\chi^2(2) = 107.181$	0.000
Time	$\chi^2(1) = 205.187$	0.000
Weekday/Weekend	$\chi^2(1) = 7.678$	0.006
Road Surface	$\chi^2(1) = 6.5$	0.011
Weather	$\chi^2(1) = 6.853$	0.009
Light Condition	$\chi^2(2) = 229.953$	0.000
Location	$\chi^2(1) = 78.846$	0.000

According to the results of Table 3, it can be identified that gender of driver, validity of license, accident cause, time, weekday/weekend, road surface, weather, light condition and location type are significantly associated with the severity. Only one variable namely alcohol test is not significantly associated with the severity. Therefore, non-significant factor is removed and continued the analysis.

One of the assumptions in logistic regression is that explanatory variables should not be highly correlated with each other. Therefore, before applying logistic regression, multicollinearity is checked among explanatory variables.

3.5. Detecting Multicollinearity between explanatory variables

The correlation coefficients among the explanatory variables can be used as first step to identify the presence of multicollinearity. Correlation matrix of highly correlated explanatory variables presented in Table 4. If Pearson correlation coefficient is greater than 0.8 or 0.9 then multicollinearity is a serious concern. Results of Table 4 indicates that the Pearson correlation coefficients between variables light condition and time as well as road surface and weather are highly correlated and indicated them as bold. These high correlation coefficients signify the presence of severe multicollinearity between the explanatory variable light condition and time of accident and road surface and weather condition.

TABLE 4: PEARSONCORRELATION MATRIX

Variables		**Time**		**Weather Condition**		**Light Condition**
		DT	*NT*	*CL*	*RA*	*GSL*
Light Condition	*DL*	**0.971** (0.000)	**-0.971** (0.000)	0.095 (0.124)	-0.095 (0.124)	**-0.837** **(0.000)**
	GSL	**-0.862** (0.000)	**0.862** (0.000)	-0.092 (0.247)	0.092 (0.247)	1.000 (0.000)
Road Surface	*D*	0.088 (0.164)	-0.088 (0.164)	**0.966** (0.000)	**-0.966** (0.000)	-0.085 (0.321)

	W	-0.088 (0.164)	0.088 (0.164)	**-0.966** (0.000)	**0.966** (0.000)	0.085 (0.326)

[a.] Cell value: correlation coefficient

[b.] P value

In addition, tolerance and VIF values were also used to confirm multicollinearity and results were indicated in Table 5.

TABLE 5: COLLINEARITY STATISTICS

Factor	**Model**	**Collinearity Statistics**	
		Tolerance	*VIF*
Validity of license	WL	.945	1.023
	WOL	.985	1.016
Time	**DT**	**.045**	**19.456**
	NT	**.050**	**20.181**
Weekend/Weekday	WE	.993	1.102
	WD	.997	1.003
Location	RD	.997	1.004
	BJ	.994	1.006
Gender	M	.994	1.006
	F	.993	1.004
Accident cause	Cause1	.971	1.030
	Cause2	.965	1.023
	Others	.959	1.043
Road surface	**D**	**.059**	**15.457**
	Others	**.067**	**14.927**
Weather	**CL**	**.061**	**15.451**
	Others	**.067**	**14.944**
Light condition	**DL**	**.052**	**19.314**
	NSL	**.057**	**18.654**
	Others	.186	5.364
Age	Age	.984	1.017

Results in Table 5 indicates very low tolerances and larger VIF values for the variables time, light condition, road surface and weather condition. Using these collinearity statistics, it can be concluded that the data almost certainly indicates a serious collinearity problem. Thus, first, variables time and road surface are removed from the data and repeat the analysis. However, collinearity still exists among the levels of light variable and weather condition. Then time and road surface variables are added and light condition and weather are removed from the analysis and continue the analysis.Results are presented in Table 6.

TABLE 6: COLLINEARITY STATISTICS OF REMAINED VARIABLES

Factor	Model	Collinearity Statistics	
		Tolerance	*VIF*
Validity of license	WL	.978	1.010
	WOL	.981	1.020
Time	DT	.980	1.028
	NT	.975	1.026
Weekend/Weekday	WE	.998	1.002
	WD	.997	1.003
Location	RD	.998	1.004
	BJ	.997	1.003
Gender	M	.993	1.007
	F	.991	1.005
Accident cause	Cause1	.639	1.565
	Cause2	.633	1.580
	Others	.638	1.560
Road surface	D	.993	1.004
	Others	.992	1.008
Age	Age	.984	1.017

According to results in Table 6, tolerances for all the predictors are very close to 1 and all the VIF values are smaller than 2.5. Therefore, it can be concluded that multicollinearity is not a concern anymore.

3.6. Checking Linearity Assumption

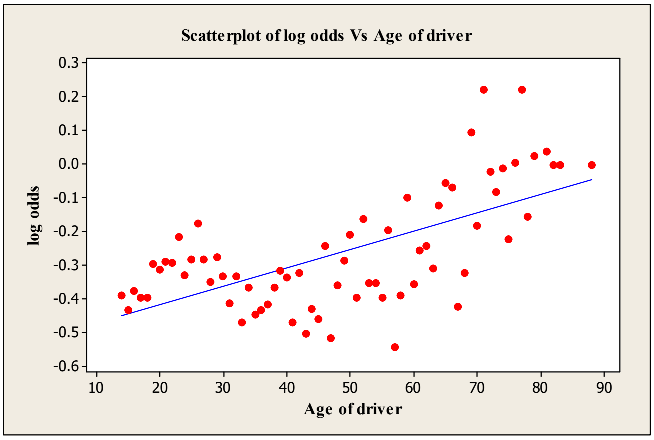

Fig 2: Log odds of severity Vs. Age of driver

The linear relationship between continuous explanatory variable age and the logit of the dependent variable is checked by visually inspecting the scatter plot and is shown in the Figure 3. It shows that variable age is linearly associated with the severity in logit scale. Thus, binary logistic regression is carried out for further analysis.

3.7. Binary logistic regression analysis

Since the response variable is dichotomous (grievous/non-grievous), the binary logistic regression model is applied for the data. The maximum likelihood procedure is used to estimate the parameters of the logistic regression model. Forward stepwise selection method was applied under the binomial logistic regression analysis, with variable entry testing based on the significance of the score statistic (the significance level was set at $p<0.05$), and removal testing based on the probability of a likelihood ratio statistic based on the maximum likelihood estimates (the significance level was set at $p>0.10$).

3.7.1. Baseline Model

Table 7 presents the results of the baseline model which is the model with only the constant included before explanatory variables are entered into the model. Logistic regression compares this model with a model including all the significant factors to determine whether the latter model is more appropriate.

TABLE 7: RESULTS OF BASELINE MODEL

Baseline Model	**B**	**S.E.**	**Wald**	**df**	**Sig.**	**Exp(B)**
Constant	.325	.014	563.373	1	.000	1.384

[c] Initial -2 Log Likelihood: 28858.35

Wald Chi-Square tests the null hypothesis that the constant equals 0. Results of the above table show that constant is statistically significant to the model. Initial log likelihood value of the baseline model is 28858.35. This value is used to select an optimal model. Predictive power of the baseline model is 56.6%, which indicates the overall percentage of correctly classified cases when there are no explanatory variables in the model.

3.7.2. Developed model interpretation

Predictors with positive coefficients cause an increasing tendency to result into fatalities. Similarly, negative coefficients indicate decreasing tendency for those significant predictors. Table 8 explains the variables in the developed model used to predict the severity of accidents.

TABLE 8: RESULTS OF BINARY LOGISTIC MODEL

Variables	**B**	**S.E.**	**Wald**	**df**	**Sig.**	**Exp(B)**	**95% CI.for EXP(B)**	
							Lower	*Upper*
Aggressive/negligent driving	.426	.028	226.060	1	.000	1.531	1.492	1.621
driving								
Night time	-.315	.029	119.342	1	.000	.730	.690	.772
Age	-.059	.030	3.914	1	.048	.942	.889	.999
Bend/Junction	-.255	.033	60.402	1	.000	.775	.726	.826
Male	.392	.055	60.394	1	.000	1.480	1.329	1.647
Constant	.592	.047	161.535	1	.000	1.807		

Results of Table 8 indicate that location type, time, age of driver, accident cause and gender have a significant effect on the severity of accidents. Time, location type and age of driver have a decreasing effect on the probability of a grievous accident. Then the rest of variables such as accident cause and gender have an increasing effect on the probability of a grievous accident. Weekday/weekend, validity of license and road surface factors are not significantly associated with the severity of accidents as p =0.218, 0.433, 0.076 >0.05 respectively. Results of the Table 8 shows that odds of a grievous accident occurred by aggressive/negligent driving is 53% more likely to be a grievous accident occurred by speeding. Odds of a grievous accident occurred in night time is 73% less likely to be a grievous accident occurred in day time. Odds of a grievous accident occurred in bend or junction is 77% less likely to be a grievous accident occurred in road. Odds of a grievous accident occurred by male motorcyclist is 48% more likely to be a grievous accident occurred by female motorcyclist. Odds ratio of age indicates that for every one unit increase in age, the odds of occurring a grievous accident decreases which implies the older the motorcyclist, the less the accident risk.

3.7.3. Logit Model

The logit model with the significant variables is:

$$\text{logit}(p) = 0.592 + 0.426\text{Cause2} - 0.315\text{NT} - 0.059\text{Age} - 0.255\text{BJ} + 0.392\text{M}$$

3.7.4. Importance of Variables in the Model

Table 9 presents the information how the model is affected if an explanatory variable is added to the model. In other words, which variable is important for the model. Results of following table are used to examine the importance of a variable in the model.

TABLE 9: IMPORTANCE OF VARIABLES IN THE MODEL

Step	**Improvement**			**Model**			**Variable**
	Chi-square	*df*	*Sig.*	*Chi-square*	*df*	*Sig.*	
1	254.230	1	.000	254.230	1	.000	IN: Accident Cause
2	119.363	1	.000	373.593	2	.000	IN: Time
3	62.783	1	.000	436.376	3	.000	IN: Location
4	60.250	2	.000	496.627	5	.000	IN: gender
5	3.910	1	.048	500.537	6	.000	IN: Age

Table 9 indicates that adding the variable accident cause to the model makes the biggest change in the model's log likelihood value. Therefore, accident cause is the most important variable in this model. It is followed by the age of driver, location type, time and gender respectively.

3.8. Measures of Goodness of Fit

3.8.1. Test of Model coefficients

TABLE 10: TEST OF DEVELOPED MODEL COEFFICIENTS

	Chi-square	**df**	**Sig.**
Step 5	3.910	1	.048
Block	500.537	6	.000
Model	500.537	6	.000

Results of Table 10 indicates that chi-square of the model is highly significant (chi-square=500.537, p=0.000 with df =6). Thus, it can be concluded that the developed model is significantly better than the baseline model. That means the accuracy of the model improved when added the explanatory variables.

3.8.2. Model Summary

TABLE 11: MODEL SUMMARY

Step	**-2 Log likelihood**	**Cox & Snell R^2**	**Nagelkerke R^2**
5	28357.808	.621	.643

According to the results of the Table 11, the developed model has a significantly reduced log likelihood value (28357.808) compared to the baseline model. It is revealed that the developed model is explaining more of the variance in the outcome and it is an improvement over the baseline model. Thus, it can be concluded that the developed model is better at predicting the severity of the accidents than the baseline model where no predictor variables were added. According to Cox & Snell R^2 and Nagelkerke R^2 values, the explained variation in the dependent variable based on the model is 62.1% and 64.3% respectively.

3.8.3. Predictive Accuracy of Developed Model

TABLE 12: CLASSIFICATION TABLE

Observed	**Predicted**		
	Grievous	*Non-grievous*	*Percentage Correct*
Grievous	5269	3940	57.21
Non-grievous	1692	11049	86.72
Overall Percentage			74.34

Table 12 indicates that 57.2% were correctly classified for grievous accidents and 86.7% for non-grievous accidents. Overall 74.3% were correctly classified. It can be seen that the developed model is correctly classifying the outcome for 74.3% of the cases compared to 56.6% in the null model.

3.8.4. Hosmer and Lemeshow test

TABLE 13: HOSMER AND LEMESHOW TEST

Step	Chi-square	df	Sig.
5	8.602	8	.377

As the results shown in the Table 13, Hosmer & Lemeshow test of the goodness of fit suggests the model is a good fit to the data as p=0.377 (>0.05).

3.8.5. Receiver Operating Characteristic Curve

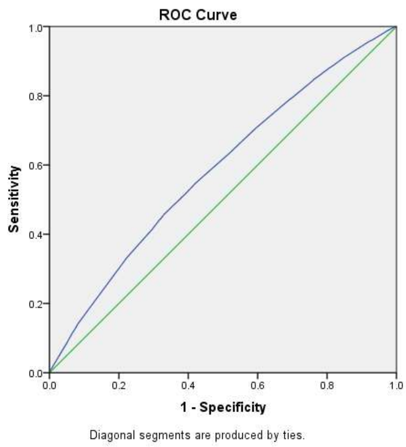

Figure 3: ROC Curve

TABLE 14: AREA UNDER THE CURVE

Area	**Std. Error**	**Asymptotic Sig.**	**Asymptotic 95% Confidence Interval**	
			Lower Bound	*Upper Bound*
.587	.004	.000	.579	.594

According to the above Table and Figure, the area under the curve is 0.587 with 95% confidence interval (0.579, 0.594). Moreover, the area under the curve is significantly different from 0.05 since the p =0.000 <0.05. That means, the logistic regression classifies the group significantly better than by chance.

3.8.6. Detecting Influential Observations

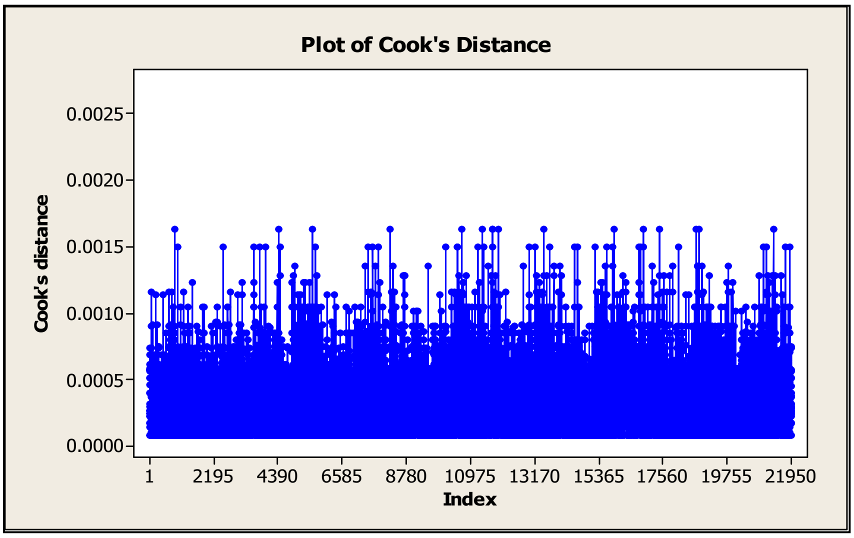

Figure 5: Plot of Cook's Distance

According to the Figure 5, it can be seen that all observations are less than 1 and even less than 0.5. Therefore, it can be said that there are neither outliers nor influential observations in this data set.

3.9. Validation of the model

TABLE 15: CATEGORY PREDICTION

Observed	**Predicted**		
	Grievous	*Non-grievous*	*Percentage Correct*
Grievous	3513	1251	73.74
Non-grievous	990	5222	84.06
Overall Percentage			79.58

Table 15 clearly indicates that the model correctly predicted 79.6% of the validation data which is greater than to the predictive power of the baseline model 56.6%. That means the developed model more accurately predicts the severity of accidents than the prediction in baseline model.

IV. CONCLUSIONS

This study was carried out to estimate the causes of motorcycles accidents in Sri Lanka. Based on the results, location type, time, age of driver, accident cause and gender have a significant effect on the severity of accidents. However, weekday/weekend, road surface and validity of license variables are not significantly associated with the severity of motor cycle accidents. Furthermore, driving on a straight road, aggressive/negligent driving, drive by male motorcyclists, drive at daytime have a high chance to be a grievous accident. Moreover, the older motorcyclists have less accident risk. These findings can aid modifying regulations and laws and establishing preventive and protective approaches and strategies.

REFERENCES

[1] Al-Ghamdi, A. (2002). Using logistic regression to estimate the influence of accident factors on accident severity. Accident Analysis and Prevention, 34, pp. 729-741.

[2] Bewick, V., Cheek, L., & Ball, J. (2005). Statistics review 14: Logistic regression. Critical Care, 9(1), 112-118.

[3] Dissanayaka, M., & Kulatunga, D. (2012). Analysis of fatalities in road accidents considering Peliyagoda police area in Gampaha. Annual Research Symposium (p. 45). University of Kelaniya.

[4] Giancristofaro, R., & Salmaso, L. (2003). Model performance analysis and model validationnin logistic regression. STATISTICA anno LXIII, 375-396.

[5] Haadi, A. (2014, May). Identification of Factors that Cause Severity of Road Accidents in Ghana: a Case Study of the Northern Region. International Journal of Applied Science and Technology, pp. 242-249.

[6] Hosmer, D., & Hjort, N. (n.d.). Goodness of fit processes for logistic regression: simulation results.

[7] Hosmer, D., Hosmer, T., Lemeshow , S., & Cessie , L. (1997). A comparison of goodness of fit tests for the logistic regression model. Statistics in medicine, 965-980.

[8] Jeepara, P., & Pirasath, S. (2011). Road traffic accidents in Eastern Sri Lanka: An analysis of admissions and outcome. Sri Lanka Journal of Surgery, 29(2), 72-76.

[9] Komba, D. D. (2006). Risk factors and road traffic accidents in Tanzania: A case study of Kibaha District. Trondheim: Norwegian University of Science and Technology.

[10] Kumarage, A. S., Wickramasingha, S. M., & Jayaratne, M. D. (2003). Analysis of road accidents in Sri Lanka. Supreme Group of Companies .

[11] Menard, S. (2002). Applied Logistic Regression Analysis (2nd ed.). A Sage University paper.

[12] Midi, H., Sarkar, S., & Rana, S. (2013). Collinearity diagnostics of binary logistic regression model. Journal of Interdisciplinary Mathematics, 253-267.

[13] Renuraj, S., Varathan, N., & Satkunananthan, N. (2015, September). Factors Influencing Traffic Accidents in Jaffna. Sri Lankan Journal of Applied Statistics, 16(2), 117-133.

[14] Senadeera, W. A. (2016). Analyzing traffic accidents in Gampaha district. University of Sri Jayewardenepura.

[15] Senaviratna, N.A.M.R. (2020). Statisyical Analysis of Severity of Motor Vehicle Accidents in Sri Lanka. Journal of Emerging Technologies and Innovative Research, 7(11), pp 481-489.

[16] Somasundaraswaran , A. K. (2006). Accident statistics in Sri Lanka. International Association of Traffic and Safety Sciences Research, 115-117.

[17] Tanaboriboon , Y., & Satinnam , T. (2005). Traffic accidents in Thailand. IATSS Research, 29(1), 88-100.

[18] Zhu, D. (2014). Analysis of factors affecting motorcycle-motor vehicle crash characteristics. Ohio: University of Dayton.