The Performance of Accident Severity Multiclass Classification

Sudesh Nair Baskara[1,2*], Haryati Yaacob[2], Sitti Asmah Hassan[2], Mohd Rosli Hainin[2]
[1]Faculty of Engineering & Quantity Surveying, INTI International University, Persiaran Perdana BBN, Putra Nilai, 71800 Nilai, Negeri Sembilan, Malaysia
[2]School of Civil Engineering, Faculty of Engineering, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia

***Email**: sudesh.baskara@newinti.edu.my

***Received**:10 January 2022; **Accepted**: 12 January 2022; **Published**:27 January 2022*

Abstract: One way to monitor accidents on highway is to analyze the accident characteristic to predict the accident severity. This study applied multinomial logistic regression model to predict accident severity. Predicted accident severities are compared with actual accident severities to evaluate the prediction performances of the model. The aim of this study is to determine the performance of accident severity classifications by multinomial logistic regression model. The predicted accident severities could be used to estimate potential effect of changes in factors contributing to accidents. Data was obtained from the Malaysian Highway Authority for the year 2013 and 2014. The accident severity was grouped into four categories of death, serious injury, minor injury and damage. Based on the results, the model correctly classified accident severities by 63.52% using training data and 61.45% using validation data. The Hosmer-Lemeshow test indicated the model has a good fit between the actual accident severities and predicted accident severities and the ROC results indicted the model able to distinguish between the classifications. The classifier of the model inclined more toward the damages compared to other accident severities resulted in classifying accident severity classes with more samples better and remains weak on the accident severity classes with lesser samples.

Keywords: Accident Severity, Model Classification, Hosmer-Lemeshow test, ROC

1. Introduction

Accident severity is one of the road safety-related aspects that requires thorough investigation. Accident severity and associated risk variables have been extensively studied and a number of studies included accident severity modelling for prediction purposes. In general, most accident-related research focused more on the results than on the performance of the model. This study evaluates and validates the performance of a multinomial logistic regression-based accident severity model with regard to pavement conditions. Accident severity model performance was determined by few researchers to analyse the model classification capability (Abdelwahab and Abdel-Aty, 2001; Ratanavaraha et al., 2014; De Oña et al., 2011) De Oña et al. (2011) used accident severity model to establish accuracy, sensitivity, and specificity. The highest accuracy was 61% accurate with ROC areas of 62%. Abdelwahab and Abdel-Aty (2001) found accuracies of 60.4% and 65.6% for training and testing sets, respectively, by modelling injury severity. Besides that, the accuracy of predicted accident severity for Rataravaraha et al. (2014) was 66.3%.

2. Methodology

This research evaluates the performances of multiclass accident severity classifications for multinomial logistic regression with the dependent variables as accident severities and the independent variables as the pavement condition factors. The accident severity data is divided into training data and validation data which are 1431 and 358 respectively. The analysis was performed with R statistical software. Four categories of the classification table which are True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) were used to assess the accuracy, sensitivity, specificity and misclassification error from the classification table. Hosmer-Lemeshow test assessed the model fit between the actual and predicted accident severities. The area under the curve (AUC) of Receiving Operating Characteristics (ROC) indicated the ability of the model to perform classification.

2.1 Classification Table

Classification table was used to classify the accident severities and predict the accuracy of the accident severity models. The classification table in Table 1 classifies both actual accident severities and the predicted accident severities. The actual accident severity numbers are the numbers derived from the available data, whereas the predicted accident severity numbers are the predictions generated by the statistical software. If the predicted accident severity value is above the threshold of 0.5, it will be considered as an event or else it will be a non-event. Referring to Table 1, True positive (TP) refers to accident severities that were observed as positive under actual classification and were predicted correctly under positive classification. True negative (TN) refers to accident severities that were observed as negative under actual classification and were predicted correctly under negative classification. False positive (FP) refers to accident severities that were observed as negative under actual classification but were predicted falsely under positive classification. False negative (FN) refers to accident severity that was observed as positive under actual classification but was predicted falsely under negative classification.

Table 1. Accident Severity Classification Table

		Actual Severity Classification	
		0 (Negative)	1 (Positive)
Predicted Severity Classification	0 (Negative)	True Negative (TN)	False Negative (FN)
	1 (Positive)	False Positive (FP)	True Positive (TP)

The accuracy of the model refers to the actual accident severity and the predicted accident severity that are correctly classified. Misclassification error is denoted as the percentage of mismatched predicted accident severities against actual accident severities. The accuracy of the classification table shown in Table 1 indicated the true labelled accident severities classification against the total true and false accident severities classification as shown in Equation 1. Misclassification error was established by the model's inaccuracy as shown in Equation 2.

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \tag{1}$$

$$\text{Misclassification Error} = (FP + FN) / (TP + TN + FP + FN) \tag{2}$$

A number of researchers have used classification table to classify accident severities. Rezapour et al. (2019) have used the severe accidents and Property Damage Only (PDO) classification table. True positive (TP) reflects the number of Property Damage Only (PDO) that is predicted to be positive. The false positives on the other hand were incidents classified as severe, notwithstanding passenger car accidents (PDO). Zhou et al. (2020) used the classification table to predict highway accidents. The agreement of observed and predicted accident conditions was true positive (TP), true negative (TN), false positive (FP) and false negatives (FN). The test results for TP and TN show the correct prediction for railway crossings while the FP and FN test show contrary predictions to the observations.

2.2 Hosmer-Lemeshow Test

The Hosmer Lemeshow Test is used to evaluate the goodness of fit of more than one outcome (Goeman and Le Cessie, 2006, Hosmer et al. 2013). This analysis aims to find out if there is evidence of model fit defined as a consensus between the actual values and predicted values (Hosmer et al. 2013). Hosmer-Lemeshow test measures the goodness of fit based on the Chi-square test. A high Chi-squared value resulted in a small p-value less than 0.05 suggests poor fit, while a low Chi-squared value with a bigger p-value closer to 1 indicates a strong model fit. When the Hosmer and Lemeshow fitness test are greater than 0.05, the null hypothesis is not rejected, showing there is no difference between the model observed and projected value. The model therefore fits the data. (Safiar et al, 2012; Jaleta, 2018)

2.3 Receiving Operating Characteristic (ROC)

The Receiving Operating Characteristics (ROC) curve is an important measure for the model's efficiency to perform classification. ROC plots true positive rate against false positive rate. Sensitivity is the amount categorised according to the real positive classes. Higher sensitivity value showed that the model is extremely reliable, since it is classified as a strong true positive value (TP). Equation 3 shows the sensitivity formula. Specificity is the test's capacity to detect true negative under the actual negative class. Equation 4 and 5 show specificity and false positive rate formulae. The area below the curve differentiates between the two accident severity groups. AUC of 0.5 to 1 indicates a good difference between the two groups whereas AUC of less than 0.5 means that the model is not able to clearly distinguish two groups.

$$\text{Sensitivity or True Positive Rate (TPR)} = [TP / (TP + FN)] \times 100 \tag{3}$$

$$\text{Specificity} = [TN / (FP + TN)] \times 100 \tag{4}$$

$$\text{False Positive Rate (FPR)} = 1 - \text{Specificity} \tag{5}$$

Chen et al. (2016) estimated that driver injury will occur in a two-year data set based on accident features. AUC of 62.7% was obtained for each level of severity of the driver injury (no injury, injury and death). Zhai et al. (2020) also constructed a real-time risk prediction model for motorways under fog circumstances. Three ROC curves have been designed based on three distinct time-scaled models, and the average fog-related accident area under the curve is 72%.

2.4 Model Validation

Validation data were used to validate the model, which represent 20% of the total data gathered, apart from the 80% of the training data. These 20% validation data have been used to test whether the model generated using training data is viable as a prediction model by checking its accuracy. The training model and validation model were utilised as an indication of comparison. A high accuracy percentage of the training model showed that the model best fit the used data, while a high accuracy percentage of the validation model showed that the model is a good predictor.

3.0 Results and Discussion

The findings in Table 2 showed actual accident severity and accident severity predicted using death, serious injury, minor injury and damage. Damage was the only appropriately predicted accident severity. The model did not predict death, serious injury, minor injury. The categorization table obtained from Table 2 is shown in Table 3. The model's accuracy was only 63.52% with a 36.48% inaccuracy. The inaccuracy was caused by the improper categorisation of major injury, minor injury and death.

Table 2. Classification Table of Training Data

		Actual Accident Severity			
	Severities	Damage	Minor Injury	Serious Injury	Death
Predicted Accident Severity	Damage	909	249	232	41
	Minor Injury	0	0	0	0
	Serious Injury	0	0	0	0
	Death	0	0	0	0

Table 3. Classification Table Output of Training Data

Accuracy	Misclassification Error	Sensitivity (TPR)	FPR	Specificity (1 - FPR)
63.52%	36.48%	100%	100%	0%

The Hosmer-Lemeshow model test produces 28.777 chi and a p-value of 0.229 > 0.05, which shows that the model well matched the actual and predicted values. The model has an average area under the curve of 0.573 (57.3%) which indicates that the model able to differentiate between classifications.

Table 4 classified the actual accident severities and predicted accident severities using validation data for death, serious injury, minor injury and damage. Similar results as the training model were seen with the validation model by referring to Table 4. Damage was correctly predicted in comparison to death, severe injury and minor injury. The output of the Table 4 classification table is shown in Table 5. The model had accuracy of 61.45% with a 38.55% inaccuracy.

Table 4. Classification Table of Validation Data

		Actual Accident Severity			
	Severities	Damage	Minor Injury	Serious Injury	Death
Predicted Accident Severity	Damage	220	68	57	13
	Minor Injury	0	0	0	0
	Serious Injury	0	0	0	0
	Death	0	0	0	0

Table 5. Classification Table Output of Validation Data

Accuracy	Misclassification Error	Sensitivity (TPR)	FPR	Specificity (1 - FPR)
61.45%	38.55%	100%	100%	0%

Table 6 presents a comparison between the training model and the validation model. The results were identical in terms of accuracy, misclassification error, sensitivity and specificity for both the training model and the validation model. This showed that the training model has been validated and the accident severity categorization is appropriately predicted.

Table 6. Classification Table Output of Training Data for The model

Data	Accuracy	Misclassification Error	Sensitivity (TPR)	FPR	Specificity (1 - FPR)
Training	63.52%	36.48%	100%	100%	0%
Validation	61.45%	38.55%	100%	100%	0%

4.0 Conclusion

The model was found to be good for predicting damage. The key reason the model ability to predict damage is because of the unbalanced data set. Multiclass data imbalance is difficult to solve due to the number of majority and minority classes is more than one (Ali et al., 2019). The data obtained recorded more accident cases with damage than minor injuries, major injury and death. The classification of the models has resulted in better accident severity classes and remains poor in the smaller severe accident classes. This makes the classifier more inclined towards damage than other accident severities. The model however resulted in an AUC above 50% indicating that the models are capable of differentiating between positive accident severity classes and negative accident severity classes. The validation model results were identical to the training model. The analyses show that the model was adequate for predicting accident severity probability without overfitting or underfitting problems.

Acknowledgements

Authors wishing to acknowledge and thank the PDRM and Malaysia Highway Authority for providing valuable information in this research.

References

Abdelwahab, H.T., & Abdel-Aty, M.A. (2001). Development of artificial neural network models to predict driver injury severity in traffic accidents at signalized intersections. Transportation Research Record. Journal of the Transportation Research Board. 1746, 6–13

Ali, H., Salleh, M. N. M., Saedudin, R., Hussain, K., & Mushtaq, M. F. (2019). Imbalance class problems in data mining: A review. Indonesian Journal of Electrical Engineering and Computer Science, 14(3), 1560-1571.

Chen, C., Zhang, G., Yang, J., Milton, J. C., & Alcántara, A. (2016). An explanatory analysis of driver injury severity in rear-end crashes using a decision table/Naïve Bayes (DTNB) hybrid classifier. Accident Analysis & Prevention, 90, 95–107.

De Oña, J., Mujalli, R. O., & Calvo, F. J. (2011). Analysis of traffic accident injury severity on Spanish rural highways using Bayesian networks. Accident Analysis & Prevention, 43(1), 402 411.

Goeman, J. J., & Le Cessie, S. (2006). A Goodness-of-Fit Test for Multinomial Logistic Regression. Biometrics, 62(4), 980–985.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied logistic regression (third ed.), John Wiley & Sons.

Jaleta, T. B. (2018). Evaluation of Risk Factors Affecting the Chance of Survival/Death Status among HIV Positive People Under the Anti Retroviral Treatment Program: The Case of Ottona Hospital. Evaluation, 57.

Ratanavaraha, V., & Suangka, S. (2014). Impacts of accident severity factors and loss values of crashes on expressways in Thailand. IATSS Research. 37(2), 130–136.

Rezapour, M., Molan, A. M., & Ksaibati, K. (2019). Analyzing injury severity of motorcycle at-fault crashes using decision tree and logistic regression methods. International Journal of Transportation Science and Technology.

Safiar, N. B., Ahmad, S., & Yacob, J. (2012). Factors Affecting Performance in Cooperative Terengganu by Using Logistic Regression. Malaysian Journal of Fundamental and Applied Sciences, 8(4).

Zhai, B., Lu, J., Wang, Y., & Wu, B. (2020). Real-time prediction of crash risk on freeways under fog conditions. International Journal of Transportation Science and Technology.

Zhou, X., Lu, P., Zheng, Z., Tolliver, D., & Keramati, A. (2020). Accident Prediction Accuracy Assessment for Highway-Rail Grade Crossings Using Random Forest Algorithm Compared with Decision Tree. Reliability Engineering & System Safety, 106931.