# FINAL REPORT WY2301F-ADDENDUM

# ASSESSMENT OF MOTORCYCLE SAFETY IN WYOMING:
# FATAL AND SEVERE CRASHES, CONTRIBUTING FACTORS AND POTENTIAL COUNTERMEASURES

Source: Milan Zlatkovic, personal archive

June 2023

Prepared by:
Milan Zlatkovic, Ph.D., P.E., PTOE
Isaac Baah

Department of Civil and Architectural Engineering and Construction Management
University of Wyoming
1000 E University Ave.
Laramie, WY 82071

**1. Report No.**
WY2301F-Addendum

**2. Government Accession No.**

**3. Recipient's Catalog No.**

**4. Title and Subtitle**
Assessment of Motorcycle Safety in Wyoming: Fatal and Severe Crashes, Contributing Factors and Potential Countermeasures - Addendum

**5. Report Date**
June 2023

6. **Performing Organization Code**

**7. Author(s)**
Milan Zlatkovic, Ph.D., P.E., PTOE, https://orcid.org/0000-0002-6777-230X
Isaac Baah, https://orcid.org/0000-0003-1596-5633
Zorica Cvijovic, Ph.D., https://orcid.org/0000-0003-3558-7967

**8. Performing Organization Report No.**
1005517

**9. Performing Organization Name and Address**
University of Wyoming
College of Engineering and Physical Sciences
1000 E University Ave
Laramie, WY 82071

**10. Work Unit No.**

**11. Contract or Grant No.**
RS04222

**12. Sponsoring Agency Name and Address**
Wyoming Department of Transportation
5300 Bishop Blvd. Bldg. 6100
Cheyenne, WY 82009

**13. Type of Report and Period Covered**
Final Report Addendum (December 2022 – June 2023)

**14. Sponsoring Agency Code**

**15. Supplementary Notes**
WYDOT Contact: Jennifer Goodrich

**16. Abstract**
Motorcycle riders and passengers are much more likely to be killed or severely injured in a crash, and on average about 15 percent of all traffic fatalities include motorcyclists. Motorcyclists and their passengers are particularly vulnerable on the roads, which accounts for their higher percentage of fatal and severe crash accidents. To take appropriate safety measures, it is important to investigate the factors that affect crash injury severity. This part of the study explores the use of two machine learning methods: Random Forest and Support Vector Machine classifiers. The Random Forest model showed better performance, therefore it was used to assess the significance of each type of contributory variable on the crash, person, and vehicle level datasets. The most important factors identified from the crash-level analysis were driver action, vehicle maneuver, type of collision, junction relation and helmet use. The driver injury area, driver action, and presence of alcohol in particular were observed to influence crash injury predictions at the person-level, whereas the vehicle maneuver, most damaged area of the vehicle and the vertical grade of the roadway were found to be relevant to motorcycle injury severity predictions at the vehicle-level. Collectively, this study revealed how motorcycle crash injury severity can be predicted from the identified factors in each data dataset. This finding will assist WYDOT and other transportation agencies in providing proactive solutions for motorcycle crashes.

**17. Key Words**
Motorcycle Safety; Crash Injury Severity; Random Forest Classifier; Support Vector Machine Classifier; Contributing Factors; Wyoming

**18. Distribution Statement**
This document is available through the National Transportation Library and the Wyoming State Library.


| 19. Security Classif. (of this report) | 20. Security Classif. (of this page) | 21. No. of Pages | 22. Price |
|---|---|---|---|
| Non-Classified | Non-Classified | 38 | |

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

# SI* (MODERN METRIC) CONVERSION FACTORS

## APPROXIMATE CONVERSIONS TO SI UNITS

| Symbol | When You Know | Multiply By | To Find | Symbol |
|---|---|---|---|---|
| | | **LENGTH** | | |
| in | inches | 25.4 | millimeters | mm |
| ft | feet | 0.305 | meters | m |
| yd | yards | 0.914 | meters | m |
| mi | miles | 1.61 | kilometers | km |
| | | **AREA** | | |
| $in^2$ | square inches | 645.2 | square millimeters | $mm^2$ |
| $ft^2$ | square feet | 0.093 | square meters | $m^2$ |
| $yd^2$ | square yard | 0.836 | square meters | $m^2$ |
| ac | acres | 0.405 | hectares | ha |
| $mi^2$ | square miles | 2.59 | square kilometers | $km^2$ |
| | | **VOLUME** | | |
| fl oz | fluid ounces | 29.57 | milliliters | mL |
| gal | gallons | 3.785 | liters | L |
| $ft^3$ | cubic feet | 0.028 | cubic meters | $m^3$ |
| $yd^3$ | cubic yards | 0.765 | cubic meters | $m^3$ |
| | | NOTE: volumes greater than 1000 L shall be shown in $m^3$ | | |
| | | **MASS** | | |
| oz | ounces | 28.35 | grams | g |
| lb | pounds | 0.454 | kilograms | kg |
| T | short tons (2000 lb) | 0.907 | megagrams (or "metric ton") | Mg (or "t") |
| | | **TEMPERATURE (exact degrees)** | | |
| °F | Fahrenheit | 5 (F-32)/9 or (F-32)/1.8 | Celsius | °C |
| | | **ILLUMINATION** | | |
| fc | foot-candles | 10.76 | lux | lx |
| fl | foot-Lamberts | 3.426 | candela/$m^2$ | cd/$m^2$ |
| | | **FORCE and PRESSURE or STRESS** | | |
| lbf | poundforce | 4.45 | newtons | N |
| lbf/$in^2$ | poundforce per square inch | 6.89 | kilopascals | kPa |

## APPROXIMATE CONVERSIONS FROM SI UNITS

| Symbol | When You Know | Multiply By | To Find | Symbol |
|---|---|---|---|---|
| | | **LENGTH** | | |
| mm | millimeters | 0.039 | inches | in |
| m | meters | 3.28 | feet | ft |
| m | meters | 1.09 | yards | yd |
| km | kilometers | 0.621 | miles | mi |
| | | **AREA** | | |
| $mm^2$ | square millimeters | 0.0016 | square inches | $in^2$ |
| $m^2$ | square meters | 10.764 | square feet | $ft^2$ |
| $m^2$ | square meters | 1.195 | square yards | $yd^2$ |
| ha | hectares | 2.47 | acres | ac |
| $km^2$ | square kilometers | 0.386 | square miles | $mi^2$ |
| | | **VOLUME** | | |
| mL | milliliters | 0.034 | fluid ounces | fl oz |
| L | liters | 0.264 | gallons | gal |
| $m^3$ | cubic meters | 35.314 | cubic feet | $ft^3$ |
| $m^3$ | cubic meters | 1.307 | cubic yards | $yd^3$ |
| | | **MASS** | | |
| g | grams | 0.035 | ounces | oz |
| kg | kilograms | 2.202 | pounds | lb |
| Mg (or "t") | megagrams (or "metric ton") | 1.103 | short tons (2000 lb) | T |
| | | **TEMPERATURE (exact degrees)** | | |
| °C | Celsius | 1.8C+32 | Fahrenheit | °F |
| | | **ILLUMINATION** | | |
| lx | lux | 0.0929 | foot-candles | fc |
| cd/$m^2$ | candela/$m^2$ | 0.2919 | foot-Lamberts | fl |
| | | **FORCE and PRESSURE or STRESS** | | |
| N | newtons | 0.225 | poundforce | lbf |
| kPa | kilopascals | 0.145 | poundforce per square inch | lbf/$in^2$ |

## Table of Contents

## List of Figures



## List of Tables

## List of Abbreviations

| | |
|---|---|
| AUC | Area under the Curve |
| CARE | Critical Analysis Reporting Environment |
| FHWA | Federal Highway Administration |
| FN | False Negatives |
| FP | False Positives |
| KABCO | Injury Severity Scale |
| MLP | Multilayer Perception |
| MNL | Multinomial Logistic Regression Models |
| NHTSA | National Highway Traffic Safety Administration |
| OOB | Out of Bag |
| RF | Random Forest |
| ROC | Receiver Operating Characteristics |
| SMOTE | Synthetic Minority Oversampling Technique |
| SVM | Support Vector Machine |
| TN | True Negatives |
| TP | True Positives |
| WYDOT | Wyoming Department of Transportation |

## Executive Summary

The popularity of motorcycles in the United States has led to an increasing emphasis on motorcycle safety. Wyoming had the highest per capita rate of motorcycle fatalities in 2019, with a rate 2.24 deaths per 100,000 population. To gain insight into the causes of these high fatality rates, two machine learning models - the Random Forest (RF) and Support Vector Machine (SVM) classifiers - were developed in this study to investigate the factors influencing motorcycle crash injury severity in Wyoming. By using the Scikit-learn package in Python, the Random Forest classifier was deemed superior after comparison. This study used data from the Wyoming Department of Transportation's (WYDOT) Critical Analysis Reporting Environment (CARE) database, spanning 12 years (2008-2019). The fatal and incapacitating injury levels were classified as 'fatal injury' (KA), while non-incapacitating, possible and no injury were grouped into the category of 'non-fatal injury' (BCO). Thus, crash severity levels were divided into two distinct categories: KA and BCO.

The data were categorized into three levels: crash, person, and vehicle, and classified based on parameters, such as environmental conditions, traffic, crash characteristics, and roadway features. Python coding language was utilized to craft and execute the machine learning algorithms. The datasets were split into training (80 percent) and test (20 percent) sets, and then the performance of both RF and SVM models were measured. Both models acquired a 76 percent correctness rate on the combined dataset; however, RF yielded a 64 percent accuracy on vehicle-level data and 58 percent on person-level data, whereas SVM registered 63 percent and 58 percent respectively. Consequently, other than the person-level data, the results revealed that RF was more precise than SVM, and was employed to identify the feature importance in terms of the contributing elements in the three datasets.

A feature importance analysis of crash data using the RF classifier revealed that the top five contributing factors to injury severity prediction were driver actions, vehicle maneuver, manner of collision, junction relation, and helmet use. When considering person-level data, injury area of the motorcyclist, driver actions, and alcohol and/or drug involvement were key elements in determining the outcome of injury severity prediction. With regards to vehicle-level data, vehicle maneuver, vehicle damage, and vertical grade of the road had a significant impact on crash injury severity prediction.

The results of the study have revealed the key influential factors that contribute to the prediction of motorcycle crash injury severity using a RF classifier. This finding is of utmost importance to WYDOT to help strategize effective means of prevention for motorcycle crashes. To further refine its insight, other approaches such as neural networks and deep learning algorithms will be utilized for the analysis of motorcycle injury severity.

## 1. Introduction

The rise in popularity of motorcycles as an alternative form of transportation has made motorcycle safety an important issue in the United States. According to the NHTSA (2023), 5,932 motorcyclists were killed in crashes throughout 2021, representing an increase of 6 percent from 2020 and about 19 percent from 2019. Furthermore, Blincoe et al. (2023) found that motorcycle rider fatalities account for around 14 percent of all motor vehicle crash deaths in the United States.

In the United States, motorcycle use is on the rise, with the number of fatal injuries from motorcycle crashes growing faster than the rate of registration. Worryingly, middle-aged riders were particularly prone to dying in such crashes (NHTSA, 2023). From 2019 to 2020, motorcycle fatalities increased by 11 percent, and 27 percent of all fatal crashes involved alcohol. Multiple studies have found that alcohol is a big factor in motorcycle injuries (Eustace et al., 2011; Jones et al., 2013; Farid and Ksaibati, 2021; Adanu et al., 2022). Moreover, Wyoming was observed to have 1.44 fatalities per 100 million vehicle miles traveled (Farid et al., 2022.). In the same year, 57 percent of motorcyclist fatalities in the state were riders who weren't wearing helmets. Interestingly, while mandatory motorcycle training is not required in Wyoming, riders aged 17 and younger are legally obligated to wear helmets (World Population Review, 2023).

In this part of the study, two machine learning models, the RF and SVM classifiers, were developed to investigate the factors influencing motorcycle crash injury severity in Wyoming. After comparing the performance of both models, the RF classifier was found to be superior, using the Scikit-learn package in Python for the analysis. To enable the study, 12 years' of statewide motorcycle crash data (2008-2019) derived from the WYDOT's Critical Analysis Reporting Environment (CARE) database were obtained. By examining several factors, including environmental, roadway, human and vehicular factors, a study by Rezapour et al. (2021) found that the two-class model worked best. Consequently, this study combined the fatal and incapacitating injury levels into the category of fatal injury (KA), while non-incapacitating, possible and no injury were grouped together as non-fatal injury (BCO). Therefore, crash severity levels were divided into two, distinct categories: KA and BCO.

Wyoming had the highest rate of motorcycle fatality per capita in 2019, with an alarming 2.24 deaths per 100,000 population according to the CARE data. To mitigate this highly disproportionate rate of fatalities, a better comprehension of the factors that lead to such high fatality must be gained. This study seeks to identify the critical factors associated with crashes, vehicles, and individuals, which can be used to develop effective interventions and thereby reduce the number of serious injuries and fatalities.

### 1.1. Study goal and methodology

The goal of this part of the study is to assess the characteristics and contributing factors of motorcycle related crashes in Wyoming through the application of machine learning algorithms. Twelve years of crash data are used to develop and compare the machine learning models for motorcycle crashes, namely RF and SVM classifiers. The data are categorized on the crash, person and vehicle level, and organized by selected variables (crash characteristics, traffic, environmental conditions and roadway characteristics). The machine learning algorithms are developed and implemented in Python programming language.

## 2. Literature Review

Machine learning has emerged as a powerful tool for improving traffic safety in recent years. This is due to the unique advantages that machine learning offers when it comes to analyzing large and complex datasets. Machine learning can process vast amounts of data and identify patterns and connections that would otherwise be too tedious or complex to identify. This can be incredibly useful for analyzing traffic safety, where data spanning weather, traffic conditions, road design, and vehicle performance all come into play.

Machine learning techniques can be utilized to analyze data from numerous sources, such as crash reports, vehicle performance tests, and traffic simulations. This analysis can then be used to develop more accurate traffic safety models. For example, one way that machine learning can be used is to develop models that can accurately predict the likelihood of a crash or collision occurring based on certain variables. Through such models, safety engineers can better determine preventive measures and safety regulations to mitigate risks.

On a broader scale, machine learning can also be used to generate insights about global trends in transportation safety. By gathering data from thousands of individual driving situations and evaluating the relationships between various variables, researchers are able to identify patterns and generate predictive models that have the ability to predict a sequence of events. Such models can be used to inform policy makers and make effective, data-driven decisions related to road safety.

In recent years, various machine learning models have been utilized in traffic safety studies, namely XGBoost, RF, SVM classifiers, Logistic Regression, deep learning models, Naïve Bayes, and K-Nearest Neighbors. In this study, RF and SVM classifiers were applied. Random Forest is a supervised learning algorithm used for regression and classification tasks (Breiman, 2001; Abdel-Aty and Haleem, 2021). It is an ensemble learning method which combines multiple decision trees together into an ensemble model. The main idea behind RF is to use many individual decision trees to make accurate predictions. It works by randomly selecting a subset of predictors for each decision tree in the forest. After selecting the predictors, a decision tree is grown. The performance of the model is improved by using a large number of decision trees, and each decision tree minimizes variance. Random Forest is both computationally efficient and extremely accurate. It is an example of an ensemble learning technique, which works by creating multiple decision trees and combining their predictions to better generalize input data. It is an effective tool for overcoming the limitations of a single decision tree, which may have over fitted a dataset. In addition, RF quickly handles large datasets with categorical and continuous variables. Random Forest has become a very popular machine learning algorithm due to its accuracy and robustness. It has also been used for a variety of applications such as demand forecasting, natural language processing, facial recognition, and image recognition.

Random Forest, first developed by Ho in 1995 and revised by Leo Breiman (Breiman, 2001), is a widely-used supervised classification algorithm that is valuable for selecting feature important variables from a collection of variables (Abdel-Aty and Haleem, 2021), employing bootstrap aggregation to reduce the high variance associated with using a single decision tree, and minimizing the misclassification error rate (Das et al., 2021). In the implementation of RF, a decision tree works well with training datasets in order to separate samples, while multiple uncorrelated decision trees are created by sub-setting the training models and making predictions via majority voting for classification (Xing et al., 2022). Furthermore, one-third of the observations are eschewed from the decision tree during tree-growing procedure, termed as out-of-bag (OOB) data points, which are utilized in computing an unbiased prediction error and estimating the measures of feature importance (Rezapour et al., 2021). In this study, Gini impurity was used for selecting the feature importance, with higher Gini node purity correlating to higher feature importance.

Random Forest is one of the models used to determine which features have an effect on the severity of a crash, and was recently utilized in a study conducted in Kentucky with deep learning models. This study was successful as the random forest achieved an accuracy of 91 percent, and revealed that factors such as collision time, crash location, driver age, and helmet use correlate with the crash severity (Xing et al., 2022). In assessing the risk factors that contribute to crash severity, the random forest classifier has shown an overall accuracy of close to 73 percent. Elyassami et al. (2020) found that disregarding traffic signals and stop signs, poor visibility, and bad weather conditions were some of the features influencing crash injury severity when they assessed the risk factors with a random forest classifier. This model was further optimized by Yan and Shen (2022) using a Bayesian optimizer. Islam et al. (2022) also used classifier models to predict road crash severity, where the random forest identified the type of collision and the cause of the crash as the main factors impacting severity.

SVM models have been widely employed in crash data analysis, and often outperform neural networks and other statistical models (Li et al., 2008). They are a type of supervised machine learning algorithm that analyzes and classifies data, capable of both linear and nonlinear classifications. By finding a subset of training data points known as SVMs, it is possible to create an optimal hyperplane that separates the classes. SVMs are common for data classification due to their efficiency and accuracy. They are particularly powerful when there are non-linear relationships between the classes or if the small training set size leads to overfitting. This decision boundary of an SVM model is drawn in such a way to have the widest possible gap between the two classes, i.e. a margin of separation. SVM models can also be used for regression. Rather than separating two categories, regression models estimate continuous range of values. Commonly, the goal of a regression model is to map the input data to the output data in a non-linear fashion. The biggest advantage of using SVM models is that they work well without risk of overfitting when data is nonlinear and are adaptable to different types of data. Moreover, the optimal solution can be found relatively quickly. However, SVM models are not applicable when the data are too large as it is computationally expensive. SVM was originally created by Boser et al. in 1992. It uses points in an N-dimensional space (here, there are only two severity levels) (Mokhtarimousavi et al., 2019) to produce an (N-1) dimensional hyperplane to separate them. SVM is then utilized to build the ideal hyperplane so that the observations can be effectively divided into groups while still maximizing the margin between the decision boundaries.

A study by Li et al. (2012) showed that the SVM model was more effective in predicting crash injury severity than an ordered probit model. Similarly, Ahmadi et al. (2020) found that the SVM model's performance was slightly better than that of multinomial and mixed multinomial logit models for predicting rear-end crashes in California. Mokhtarimousavi et al. (2019) used a mixed logit model and an SVM with metaheuristic algorithm enhancements, finding that the SVM provided higher predictive accuracy than the traditional statistical model. Yu and Abdel-Aty (2013) then compared the accuracy of an SVM model with various kernels to that of a Bayesian logistic model and found that the SVM with radial-basis kernel performed the best. Sharma et al. (2016) also compared the SVM model (using a Gaussian Kernel) to a Multilayer Perception (MLP) for crash prediction and observed that the SVM achieved an accuracy of 94 percent on the test data, while the MLP only achieved an accuracy of 60 percent. Lastly, Aghayan et al. (2015) conducted a comparison of several models (namely, SVM, MLP, genetic algorithm, combined genetic algorithm, and pattern search) and determined that the MLP achieved an accuracy of 86.2 percent, outperforming the SVM model (with an overall accuracy of 81.4 percent). Despite its advantages, the SVM model has some drawbacks, the main being that it operates as a black box and does not possess a functional form between crashes and covariates, and that it requires three parameters be determined before training (Li et al., 2008).

## 3. Methodology

For this study, crash data ranging from 2008 – 2019 on crash, vehicle and person levels were taken from the WYDOTCARE database. This subset included a range of environmental, roadway, human and vehicle factors. Crash severity levels in the database were classified into Fatal (K), Incapacitating (A), Non-Incapacitating (B), Possible (C) and No Injury (O). However, for the purpose of this study, the two categories of 'Fatal and Incapacitating' and 'Non-Incapacitating, Possible and No Injury' were consolidated to form binary categories of 'Fatal Injury' (KA) and 'Non-Fatal Injury' (BCO). Three sets of data were prepared and used in this study. The crash data comprised of 21 explanative variables, while personal and vehicle data both covered six explanative variables.

A total of 3,127 motorcycle-related crashes were extracted during the pre-processing phase, with 2,050 of the records being non-fatal and 1,077 being fatal. The categorical data types were then converted to binary variables (1 and 0) using the one-hot encoding method in Python, to indicate the presence or absence of a certain class. Since the data were found to be imbalanced, with 2,050 non-fatal and 1,077 fatal records, the Synthetic Minority Oversampling Technique (SMOTE) was utilized to balance the dataset. Once SMOTE was applied, an equal number of fatal and non-fatal crash records were generated. Finally, the data were randomly split into training and testing datasets, with 80 percent of the data used as the training datasets, and 20 percent of the data used to validate the model.

In the analysis, RF and SVM machine models were applied to the crash data. The models were implemented in Python using Scikit-learn, which is a popular package for machine learning. The performance of machine learning algorithms is evaluated using multiple metrics, including classification accuracy, precision, recall, f-measure, and area under the receiver operating characteristics curve. The factored confusion matrix models these metrics with four sections, true-positives (TP), true-negatives (TN), false-positives (FP), and false-negatives (FN), with the following definitions:

- TP indicates that, when the crash severity was fatal, the model predicted fatal.
- TN indicates that, when the crash severity was non-fatal, the model predicted non-fatal.
- FP indicates that, when the crash severity was fatal, the model predicted non-fatal.
- FN indicates that, when the crash severity was non-fatal, the model predicted fatal.

The Area under the Curve (AUC) is a metric used to assess the accuracy of machine learning classifiers. The receiver operating characteristics (ROC) curve, represented as an AUC value, is constructed by plotting the true positive rate on the y-axis and the false positive rate on the x-axis (Huang and Ling, 2005). According to McDowell (2006), AUC values between 0.5 and 0.7 indicate low accuracy, ones ranging from 0.7 to 0.9 represent moderate accuracy, and values higher than 0.9 signify high accuracy.

## 4. Results

The dataset used in this study was split in such a way that 80 percent of data were being designated for training, and 20 percent for testing. The test data consisted of 820 observations. The injury severity was considered binary, with fatal labeled as 1 and non-fatal labeled as 0. Sixteen, six, and six independent variables or features pertaining to the crash, person, and vehicle levels respectively, were included in the analysis to construct models in order to prognosticate motorcycle crash injury severity. Multiple models were developed, with RF and SVM machine models both chosen based on the accuracy of their models. In evaluating these models, confusion matrices and ROC-AUC were both utilized.

Table 1 displays the metrics used to compare the two models: recall/ sensitivity, precision, and F1-score. A model can have a higher recall and lower precision, or vice versa. The F1-score, or F-measure, is the measure used to ensure the precision and recall are balanced, and a higher F1-score is an indicator of a model that is balanced with both high recall and precision values. The SVM model achieved a total accuracy of 73 percent, with an F1-score of 72 percent and 73 percent for fatal and non-fatal, respectively. Its precision was 74 percent and 72 percent, and its recall/ sensitivity was 71 percent and 74 percent for fatal and non-fatal severity levels. The RF classifier, on the other hand, achieved a better overall accuracy than SVM, with 76 percent overall accuracy and an F1-score of 77 percent and 74 percent for fatal and non-fatal respectively. It achieved precision values of 78 percent and 74 percent, and recall/ sensitivity values of 81 percent and 71 percent for fatal and non-fatal severity levels respectively for the crash level.

For the person level, the accuracy of the RF and SVM Classifiers was the same at 58 percent, however the performance of SVM was more uniform than that of RF. In the vehicle level, the RF achieved a slightly higher accuracy of 64 percent, while the SVM produced 63 percent. The SVM F1-score was 61 percent and 65 percent for fatal and non-fatal respectively, with precision metric of 65 percent and 61 percent, and recall (or sensitivity) of 57 percent and 68 percent. The RF F1-score was 62 percent for both fatal and non-fatal, with precision of 65 percent and 62 percent, and recall of 59 percent and 68 percent respectively. Precision metric measures the exactness of the model's results, while recall (or sensitivity) looks at the ratio of all positively-labelled examples to the total truly positive, respectively.

The ROC-AUC is another metric often used for assessing the performance of machine learning classifiers. The ROC is the probability, and AUC is a measure of separability. The ROC-AUC metric gauges the performance of the models, with a higher ROC-AUC indicating a better capability for the model to distinguish between the classes in question (in this instance, fatal and non-fatal). Both algorithms appear to perform well for the given dataset, however the RF

Classifier achieved a slightly higher AUC score than the SVM Classifier, except for the person-level data.

**Table 1. Random Forest and SVM Classifier Performance Measures.**

| Levels | Model | Crash Severity Level | Recall/Sensitivity (%) | Precision (%) | F1-Score (%) |
|---|---|---|---|---|---|
| Crash Level | Random Forest Classifier | Fatal | 81 | 78 | 77 |
| | | Non-Fatal | 71 | 74 | 74 |
| | | Overall Accuracy (%) | 76 | | |
| | SVM Classifier | Fatal | 71 | 74 | 72 |
| | | Non-Fatal | 74 | 72 | 73 |
| | | Overall Accuracy (%) | 73 | | |
| | | **Crash Severity Level** | **Recall/Sensitivity (%)** | **Precision (%)** | **F1-Score (%)** |
| Person Level | Random Forest Classifier | Fatal | 73 | 56 | 64 |
| | | Non-Fatal | 43 | 61 | 51 |
| | | Overall Accuracy (%) | 58 | | |
| | SVM Classifier | Fatal | 58 | 58 | 58 |
| | | Non-Fatal | 57 | 58 | 57 |
| | | Overall Accuracy (%) | 58 | | |
| | | **Crash Severity Level** | **Recall/Sensitivity (%)** | **Precision (%)** | **F1-Score (%)** |
| Vehicle Level | Random Forest Classifier | Fatal | 59 | 65 | 62 |
| | | Non-Fatal | 68 | 62 | 62 |
| | | Overall Accuracy (%) | 64 | | |
| | SVM Classifier | Fatal | 57 | 65 | 61 |
| | | Non-Fatal | 68 | 61 | 65 |
| | | Overall Accuracy (%) | 63 | | |

## 5. Discussion

Due to the fact that the RF classifier performed better overall on the analyzed dataset, it was to assess the importance of crash contributing factors. A value is assigned to each feature to establish its importance in predicting crash severity. The feature importance was assessed separately on the crash, person and vehicle levels.

### 5.1. Crash level feature importance

Figure 1 shows the feature importance plot for crash-level data. A mean decrease in Gini score revealed that driver actions had the greatest effect on crash severity. Such actions as running off the road, failing to keep in the right lane, erratic or aggressive driving, other improper actions, avoiding moving vehicles, and failing to yield the right of way all had a direct influence on the level of motorcycle crash severity.

Vehicle maneuvers are the second biggest factor contribution to crash severity. A large number of crashes occur when the rider is negotiating a curve. Additionally, the type of impact (angle same direction, front-to-side, rear-end, front-to-rear angle front-to-side opposing direction) is another factor that affects injury severity. Location is also influential, with crashes happening at non-junction intersections often leading to greater severity of rider injuries. Wearing a helmet is of vital importance as well, as those involved in crashes without one tend to result in more severe consequences. Weekday and weekends, lighting conditions, seasonal variations, and the presence of wild animals could also affect crash severity prediction. Finally, road conditions, weather, drug/alcohol suspicions, and speeding appear to be the least significant factors.

### 5.2. Person level feature importance

Figure 2 shows the feature importance of the RF Classifier on the person level data. The results indicate that driver injury area is the leading factor in predicting the severity of motorcycle crash injury, with particular attention to the upper extremities (arms, hands, shoulders), followed by head, lower extremities (legs, feet, etc.), and thorax (chest/back). Driver age was found to be the second most influential factor in this prediction, with young riders having a greater impact than middle and older-age riders. The least impactful factors were drugs involved and helmet use.

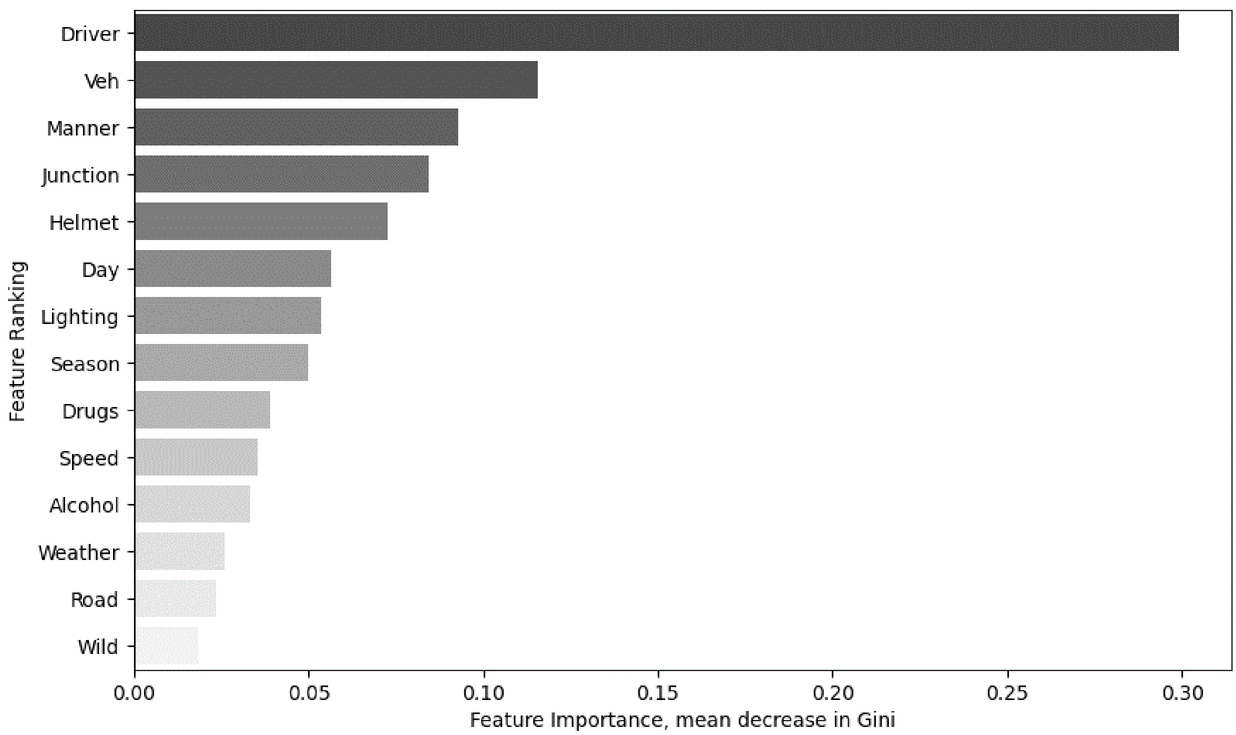


**Figure 1. Diagram. Random Forest relative feature importance for crash level data.**

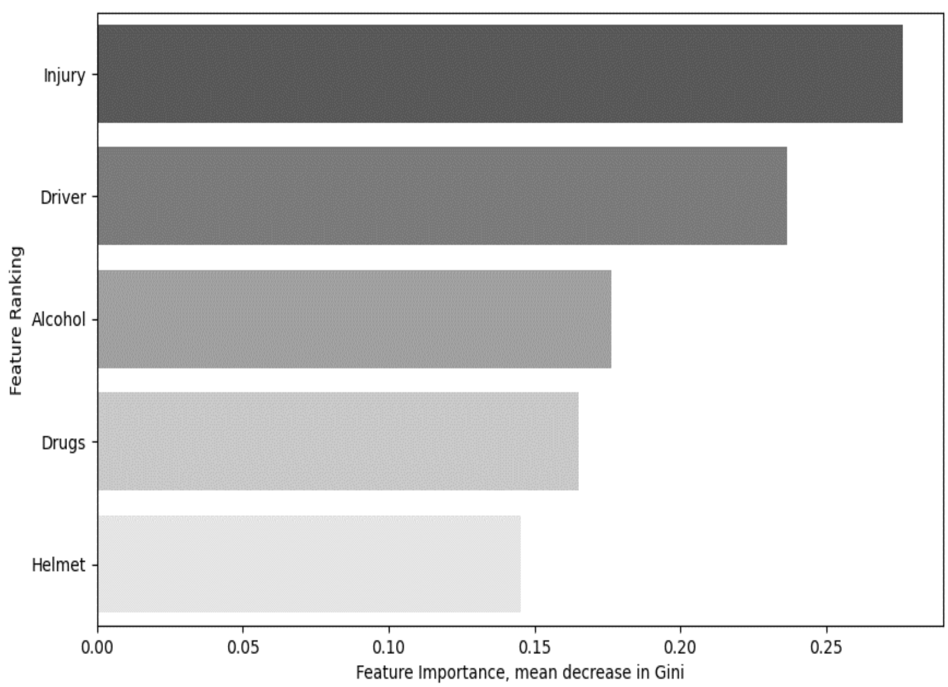


**Figure 2. Diagram. Random Forest relative feature importance for person level data.**

### 5.3. Vehicle level feature importance

Figure 3 shows the feature importance ranking of the RF Classifier for the vehicle-level data. Negotiating curves, travelling in a straight line and making right turns have been identified as major contributory factors to crash injury severity when it comes to motorcycle maneuver. Additionally, vehicle damage, particularly when disabling or functional, is a prominent predictor of crash injury severity. Moreover, the vertical grade of the roadway has been found to have an influence on the crash severity outcome. The manner of collision and the horizontal alignment of the roadway are found to be the least influential in terms of crash injury prediction.

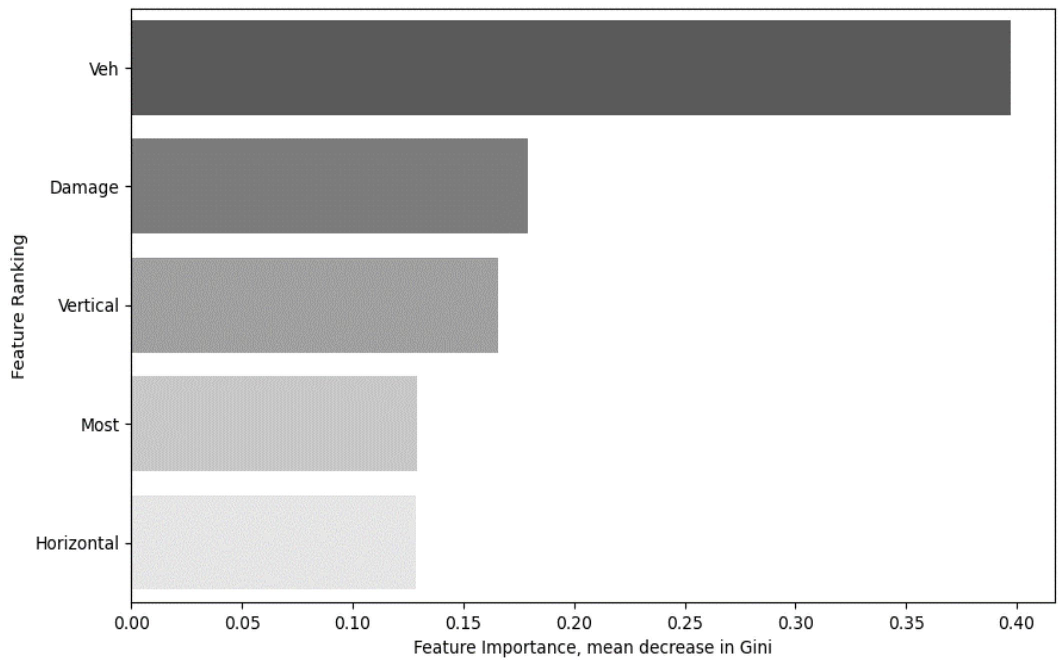


**Figure 3. Diagram. Random Forest relative feature importance for vehicle level data.**

## 6. Comparison of Machine Learning and Statistical Models

This section compares the crash contributing factors found in the predictive models of machine learning and statistical models used to analyze motorcycle crash injury severity in Wyoming. The statistical models utilized for this research included multinomial logit and Bayesian regression models, conducted in the preceding stage of the study. Additionally, the analysis was done on crash, person, and vehicle levels, allowing for the comparison between the machine learning models based on the same categorization.

### 6.1. Crash level analyses comparison

The results of the RF classifier on the crash level are provided in Table 2. Based on the feature impact analysis of the RF classifier, it was found that when the roadway surface condition is dry, the probability of predicting motorcycle fatal/incapacitating injury severity is higher compared to other conditions such as wet, ice, dirt, gravel, etc. The multinomial logit (MNL) for the rural single model showed that most roadways conditions, excluding ice or frost, decrease the odds of fatal injury. Cloudy or overcast fog and severe wind were identified as increasing the odds of fatal injury for rural single crashes. The RF classifier showed a higher probability for predicting fatal/incapacitating injury severity when the weather condition is clear when compared to the other included conditions. Regarding vehicle maneuver, straight ahead movement increased the probability of predicting fatal/incapacitating injury severity higher than other vehicle maneuvers. Excluding entering a traffic lane, making a U-turn, negotiating a curve, and overtaking or passing, most of the vehicle maneuvers were found to lower the odds of fatal injury for the MNL model.

When alcohol, an animal or speeding is involved, the odds of fatal injury for a rural or urban single-vehicle MNL model increased significantly. In contrast, the RF model found that the probability of fatal/incapacitating injury decreased when compared to a situation where no alcohol is involved in the crash. No helmet use was found to increase the odds of fatal injury for rural and urban single and multi-vehicle of the MNL model. Moreover, the RF model revealed that not wearing a helmet increased the probability of predicting fatal/incapacitating injury by more than 45 percent, when compared to helmet use.

Intersections, private roads, and through roadways at junction relations were found to increase the odds of fatal injury; however, the odds decreased for business entrances, driveways, and ramps. The RF model revealed that when the junction relation is non-junction, the probability of predicting fatal/incapacitating injuries becomes higher in comparison to other junction relations. Alcohol and speeding were found to raise the odds of fatal injury for rural and urban multi-vehicle MNL model, whereas the RF model showed that the probability of fatal/incapacitating injuries decreased when no alcohol was present. The MNL model revealed that no helmet use decreased the odds of fatal injury, whereas the RF model demonstrated that no helmet use increased the probability of predicting fatal/incapacitating injuries. In addition, no improper driver action was found to significantly increase the probability of predicting fatal/incapacitating injury when compared to other driver actions. Finally, for the urban single MNL model, the driver actions such as disregarding other road markings, evading law enforcement, improper passing, and speeding were found to drastically increase the odds of fatal injury.

The MNL model found that all types of multi-vehicle collisions, with the exception of rear-to-front backing, can increase the likelihood of severe crash outcomes for motorcycles. Conversely, the RF model demonstrated that the lack of collisions involving two vehicles improves the predictability of fatal/incapacitating injury when compared to the other forms of collisions.

**Table 2. Random Forest Feature Impact Analysis on Crash Level.**

| Variables | Severity Levels | |
|---|---|---|
| | Fatal/Incapacitating (KA) | Non-Incapacitating/Possible/ No Injury (BCO) |
| ***Alcohol Involved*** | | |
| No | 0.905 | 0.784 |
| Yes | 0.095 | 0.216 |
| ***Driver Action*** | | |
| Avoiding an Object on the Road | 0.007 | 0.000 |
| Avoiding Animal | 0.027 | 0.007 |
| Avoiding MV | 0.019 | 0.026 |
| Avoiding Non-Motorist | 0.000 | 0.000 |
| Disregarded Other Road Marking | 0.005 | 0.011 |
| Disregarded Traffic Signs | 0.010 | 0.017 |
| Drove too Fast for Conditions | 0.036 | 0.078 |
| Erratic or Reckless or Careless or aggressive | 0.036 | 0.052 |
| Evading Law Enforcement | 0.010 | 0.005 |
| Failed to Keep Proper Lane | 0.063 | 0.091 |
| Failed to Yield ROW | 0.056 | 0.090 |
| Following too Close | 0.083 | 0.048 |
| Improper Backing | 0.007 | 0.000 |
| Improper Passing | 0.005 | 0.030 |
| Improper Turn or No Signal | 0.024 | 0.013 |
| No Improper Driving | 0.369 | 0.214 |
| Other Improper Action | 0.029 | 0.027 |
| Over Corrected or Over Steered | 0.012 | 0.014 |
| Ran Off Road | 0.080 | 0.142 |
| Ran Red Light | 0.005 | 0.000 |
| Speeding | 0.044 | 0.070 |
| Swerve Due to Wind or Slippery Surface | 0.015 | 0.007 |
| Wrong Side or Wrong Way | 0.000 | 0.004 |
| ***Helmet*** | | |
| Helmet Used | 0.381 | 0.355 |
| None Used | 0.553 | 0.607 |

**Table 2 Continued**

| | | |
|---|---|---|
| ***Junction Relation*** | | |
| Business Entrance | 0.056 | 0.029 |
| Crossover Related | 0.000 | 0.000 |
| Driveway Related | 0.029 | 0.023 |
| Entrance or Exit Ramp | 0.000 | 0.000 |
| Intersection | 0.362 | 0.291 |
| Non-Junction | 0.493 | 0.616 |
| Other Non-Interchange eg., Trail or School Xing | 0.010 | 0.006 |
| Other Parts eg., Gore | 0.017 | 0.003 |
| Thru Roadway | 0.019 | 0.020 |
| ***Lighting Condition*** | | |
| Darkness Lighted | 0.075 | 0.047 |
| Darkness Unlighted | 0.087 | 0.110 |
| Dawn | 0.015 | 0.002 |
| Daylight | 0.799 | 0.814 |
| Dusk | 0.024 | 0.025 |
| ***Manner of Collision*** | | |
| Angle Front to Side Opposing Direction | 0.022 | 0.053 |
| Angle Right Front to Side includes Broadside | 0.107 | 0.097 |
| Angle Same Direction Front to Side | 0.056 | 0.030 |
| Head On Front to Front | 0.015 | 0.040 |
| Not a Collision with 2 Vehicles in Transport | 0.614 | 0.658 |
| Other | 0.012 | 0.002 |
| Rear End Front to Rear | 0.133 | 0.093 |
| Rear to Front Normally Backing | 0.002 | 0.000 |
| Rear to Rear Normally Backing | 0.000 | 0.000 |
| Rear to Side Normally Backing | 0.002 | 0.000 |
| Sideswipe Opposite Direction Meeting | 0.012 | 0.013 |
| Sideswipe Same Direction Passing | 0.024 | 0.010 |
| ***Road Condition*** | | |
| Dry | 0.896 | 0.946 |
| Ice or Frost | 0.005 | 0.000 |
| Dirt or Gravel | 0.027 | 0.022 |
| Oil or Fuel | 0.007 | 0.000 |
| Sand on Dry Pavement | 0.015 | 0.002 |
| Wet | 0.044 | 0.020 |

**Table 2 Continued**

| | | |
|---|---|---|
| ***Speeding*** | | |
| No | 0.852 | 0.724 |
| Yes | 0.148 | 0.276 |
| ***Vehicle Maneuver*** | | |
| Backing | 0.002 | 0.000 |
| Changing Lanes | 0.012 | 0.012 |
| Entering a Traffic Lane | 0.007 | 0.015 |
| Leaving a Traffic Lane | 0.000 | 0.005 |
| Making a U-Turn | 0.007 | 0.002 |
| Negotiating a Curve | 0.167 | 0.281 |
| Other | 0.007 | 0.005 |
| Overtaking or Passing | 0.029 | 0.064 |
| Parked | 0.002 | 0.000 |
| Slowing | 0.058 | 0.025 |
| Stopped in Traffic | 0.010 | 0.003 |
| Straight Ahead | 0.527 | 0.477 |
| Turning Left | 0.102 | 0.092 |
| Turning Right | 0.058 | 0.014 |
| ***Weather*** | | |
| Clear | 0.893 | 0.922 |
| Cloudy or Overcast | 0.049 | 0.044 |
| Fog | 0.000 | 0.000 |
| Raining | 0.044 | 0.018 |
| Severe Wind Only | 0.005 | 0.012 |
| Sleet, Hail, or Freezing Rain | 0.000 | 0.000 |
| Snowing | 0.002 | 0.000 |
| ***Animal Involved*** | | |
| No | 0.949 | 0.922 |
| Yes | 0.051 | 0.078 |

### 6.2. Vehicle level analyses comparison

Based on the feature impact analysis of the RF classifier, provided in Table 3, it was found that a straight horizontal alignment increases the probability of predicting fatal/incapacitating injuries in comparison to curving left and/or right. Furthermore, MNL model showed that curving left and/or right increases the odds of fatal injury. Additionally, when the vertical grade has a level grade, it increases the probability of fatal/incapacitating injury. In contrast, the MNL models showed that when the vertical grade is crest or sag, the odds of fatal injury are higher; on the other hand, downhill and uphill decrease the odds of fatal injury. Moreover, the vehicle damage that is disabling, functional, or minor all profoundly influence the prediction of fatal/incapacitating crashes compared to having no vehicle damage, which is consistent with the MNL model.

Vehicles that are older than 10 years increase the chances of predicting fatal or incapacitating injury when compared to those that are five years old or less, or between six to 10 years old. The MNL model also indicated that when vehicle age is greater than 10 years, the odds of fatal injury rise. Collisions that involve any part of the vehicle body pose a higher likelihood of fatal/incapacitating injury when compared to incidents without collision. The model further discovered that most vehicle body parts (left front area, left side, right front area, etc.) increase the odds of fatal injury. According to findings, when the vehicle maneuver is straight, the probability of fatal/incapacitating injury is higher compared to other maneuvers, with the exception of overtaking/passing, parked, and slowing.

### 6.3. Person level analyses comparison

The MNL model suggested that, on the person level, the odds of a fatal injury increased when the driver involved was middle-aged. Conversely, the RF model results, given in Table 4, revealed that, when a crash involved an elderly rider, the probability of fatal or incapacitating injury was 3.7 percent and 31.5 percent higher than for middle-aged and young drivers, respectively. With regards to gender, the random forest model revealed that, when a male driver was involved in a crash, the probability of fatal or incapacitating injury was higher than for female drivers. The MNL model, meanwhile, found that male drivers increased the odds of fatal injury whilst female drivers decreased them. Finally, when a crash involved a driver not wearing a helmet, the random forest model showed an increased probability of fatal or incapacitating injury.

The MNL model showed that wearing a helmet significantly reduces the odds of fatal injury. Interestingly, the MNL model also showed that when alcohol or drugs are involved in a crash, the odds of fatal injury increase. In contrast, the random forest model found that when alcohol or drugs are involved in a crash, the probability of a fatal or incapacitating injury decreases. Furthermore, when compared to other injury areas, the MNL model showed that having no injury increases the odds of fatal injury. All other injury areas were found to increase the odds of fatal injury.

**Table 3. Random Forest Feature Impact Analysis on Vehicle Level.**

| Variables | Severity Levels | |
|---|---|---|
| | Fatal/Incapacitating (KA) | Non-Incapacitating/Possible/No Injury (BCO) |
| ***Horizontal Alignment*** | | |
| Curve Left | 0.068 | 0.063 |
| Curve Right | 0.046 | 0.058 |
| Straight | 0.830 | 0.815 |
| ***Vertical Grade*** | | |
| Downhill | 0.126 | 0.087 |
| Hillcrest | 0.005 | 0.010 |
| Level | 0.735 | 0.759 |
| Sag | 0.005 | 0.010 |
| Uphill | 0.083 | 0.074 |
| ***Vehicle Damage*** | | |
| Disabling | 0.320 | 0.353 |
| Functional | 0.262 | 0.235 |
| Minor | 0.286 | 0.265 |
| None | 0.049 | 0.061 |
| ***Vehicle Year*** | | |
| Between 6-10 | 0.036 | 0.038 |
| Greater than 10 | 0.893 | 0.908 |
| Less or equal to 5 | 0.046 | 0.034 |
| ***Most Damaged Area*** | | |
| Non-Collision | 0.061 | 0.063 |
| Top | 0.017 | 0.009 |
| Undercarriage | 0.027 | 0.022 |
| Vehicle Body | 0.799 | 0.768 |
| ***Vehicle Maneuver*** | | |
| Changing Lanes | 0.012 | 0.007 |
| Entering a Traffic Lane | 0.007 | 0.010 |
| Leaving a Traffic Lane | 0.000 | 0.007 |
| Making a U-Turn | 0.007 | 0.004 |
| Negotiating a Curve | 0.167 | 0.318 |
| Other | 0.007 | 0.004 |
| Overtaking or Passing | 0.029 | 0.055 |
| Parked | 0.002 | 0.000 |
| Slowing | 0.058 | 0.030 |
| Stopped in Traffic | 0.010 | 0.001 |
| Straight Ahead | 0.527 | 0.442 |
| Turning Left | 0.102 | 0.103 |
| Turning Right | 0.058 | 0.016 |

**Table 4. Random Forest Feature Impact Analysis on Person Level.**

| Variables | Severity Levels | |
|---|---|---|
| | Fatal/Incapacitating (KA) | Non-Incapacitating/Possible/No Injury (BCO) |
| ***Driver Age*** | | |
| Middle | 0.354 | 0.385 |
| Old | 0.367 | 0.428 |
| Young | 0.279 | 0.183 |
| ***Drivers Gender*** | | |
| Female | 0.148 | 0.142 |
| Male | 0.840 | 0.856 |
| ***Alcohol Involved*** | | |
| No | 0.905 | 0.815 |
| Yes | 0.095 | 0.185 |
| ***Drugs Involved*** | | |
| No | 0.930 | 0.834 |
| Yes | 0.012 | 0.043 |
| ***Helmet*** | | |
| Helmet Used | 0.381 | 0.355 |
| None Used | 0.553 | 0.612 |
| ***Injury Area*** | | |
| Abdomen/Pelvis | 0.017 | 0.004 |
| Face | 0.032 | 0.017 |
| Head | 0.049 | 0.098 |
| Lower Extremities (Legs, Feet, etc.) | 0.041 | 0.059 |
| No Injury | 0.609 | 0.630 |
| Spine | 0.015 | 0.005 |
| Thorax (Chest/Back) | 0.024 | 0.041 |
| Upper Extremities (Arms, Hand, Shoulder) | 0.056 | 0.043 |

## 7. Conclusions

This part of the study employed machine learning methods, RF and SVM classifiers, to investigate the contributing factors to motorcycle injury severity. Data were sourced from WYDOT's CARE database, and comprised of motorcycle crash data spanning from 2008 to 2019. The injury severity levels recorded in the CARE database were divided into two categories (KA and BCO) based on previous research findings in the area. Subsequently, the datasets were split into training (80 percent) and test (20 percent) subsets, and both RF and SVM classifier accuracy was evaluated. Both achieved a 76 percent accuracy on the overall dataset; however, RF reached 64 percent accuracy for the vehicle-level data and 58 percent for the person-level data whereas SVM recorded 63 percent and 58 percent, respectively. As a result, apart from person-level data, the results indicated that RF was more accurate than the SVM, and was employed to identify the feature importance regarding the contributing factors in the three datasets.

Using the RF classier, a feature importance analysis of crash data showed that driver actions, vehicle maneuver, manner of collision, junction relation, and helmet use were the top five contributing factors to the prediction of crash injury severity. For person-level data, injury area of the motorcyclist, driver actions, and involvement of alcohol and/or drugs were key factors in determining the outcome of injury severity prediction. For vehicle-level data, vehicle maneuver, vehicle damage, and vertical grade of the road were the features that influenced the prediction of crash injury severity.

Results of the study have revealed the most influential factors that are contributing to the prediction of motorcycle crash injury severity using the RF classifier. This result will be invaluable to WYDOT, as it will help them to develop a proactive solution to motorcycle crashes. As next steps, other approaches such as neural networks and deep learning algorithms will be employed to analyze motorcycle injury severity.

## References


1. Abdel-Aty, M., and Haleem, K. (2011). Analyzing angle crashes at unsignalized intersections using machine learning techniques. *Accident Analysis and Prevention*, *43*(1), 461–470. https://doi.org/10.1016/j.aap.2010.10.002.

2. Adanu, E. K., Agyemang, W., Islam, R., and Jones, S. (2022). A comprehensive analysis of factors that influence interstate highway crash severity in Alabama. *Journal of Transportation Safety and Security*, *14*(9), 1552–1576. https://doi.org/10.1080/19439962.2021.1949414.

3. Aghayan, I., Hosseinlou, H., and Kunt, M. M. (2015). Application of Support Vector Machine for Crash Injury Severity Prediction: A Model Comparison Approach. In *Journal of Civil Engineering and Urbanism* (Vol. 5, Issue 5).

4. Ahmadi, A., Jahangiri, A., Berardi, V., and Machiani, S. G. (2020). Crash severity analysis of rear-end crashes in California using statistical and machine learning classification methods. *Journal of Transportation Safety and Security*, *12*(4), 522–546. https://doi.org/10.1080/19439962.2018.1505793.

5. Blincoe, L., Miller, T. R., Wang, J.-S., Swedler, D., Coughlin, T., Lawrence, B., Guo, F., Klauer, S., and Dingus, T. (2023). *The Economic and Societal Impact of Motor Vehicle Crashes, 2019*. NHTSA, Report No. DOT HS 813 403.

6. Boser, B. E., Guyon, I. M., and Vapnik, V. N. (1992). *A Training Algorithm for Optimal Margin Classifiers*. COLT '92: Proceedings of the fifth annual workshop on Computational learning theory, pp. 144–152. https://doi.org/10.1145/130385.130401

7. Breiman, L. (2001). *Random Forests*. Machine Learning Vol. 45, pp. 5–32. https://doi.org/10.1023/A:1010933404324

8. Das, S., Datta, S., Zubaidi, H. A., and Obaid, I. A. (2021). Applying interpretable machine learning to classify tree and utility pole-related crash injury types. *IATSS Research*, *45*(3), 310–316. https://doi.org/10.1016/j.iatssr.2021.01.001.

9. Elyassami, S., Hamid, Y., and Habuza, T. (2020). Road crash analysis and prediction using gradient-boosted and random forest trees. *Colloquium in Information Science and Technology, CIST*, *2020-June*, 520–525. https://doi.org/10.1109/CiSt49399.2021.9357298.

10. Eustace, D., Indupuru, V. K., and Hovey, P. (2011). Identification of risk factors associated with motorcycle-related fatalities in Ohio. *Journal of Transportation Engineering*, 137(7), 474–480. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000229

11. Farid, A., and Ksaibati, K. (2021). Modeling severities of motorcycle crashes using random parameters. *Journal of Traffic and Transportation Engineering (English Edition)*, *8*(2), 225–236. https://doi.org/10.1016/j.jtte.2020.01.001.

12. Farid, A., Nazneen, S., Ksaibati, K., and Rezapour, M. M. (n.d.). *Enhancing Crash Data Reporting to Highway Safety Partners in Wyoming by Utilizing Big Data Analysis and Survey Techniques (MPC-22-482)*.

13. Huang, J., and Ling, C. X. (2005). Using AUC and accuracy in evaluating learning algorithms. *IEEE Transactions on Knowledge and Data Engineering*, *17*(3), 299–310. https://doi.org/10.1109/TKDE.2005.50.

14. Islam, M. K., Reza, I., Gazder, U., Akter, R., Arifuzzaman, M., and Rahman, M. M. (2022). Predicting Road Crash Severity Using Classifier Models and Crash Hotspots. *Applied Sciences (Switzerland)*, *12*(22). https://doi.org/10.3390/app122211354.

15. Jones, S., Gurupackiam, S., and Walsh, J. (2013). Factors influencing the severity of crashes caused by motorcyclists: Analysis of data from Alabama. *Journal of Transportation Engineering*, 139(9), 949–956. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000570

16. Li, X., Lord, D., Zhang, Y., and Xie, Y. (2008). Predicting motor vehicle crashes using Support Vector Machine models. *Accident Analysis and Prevention*, *40*(4), 1611–1618. https://doi.org/10.1016/j.aap.2008.04.010.

17. Li, Z., Liu, P., Wang, W., and Xu, C. (2012). Using support vector machine models for crash injury severity analysis. *Accident Analysis and Prevention*, *45*, 478–486. https://doi.org/10.1016/j.aap.2011.08.016.

18. Mcdowell, I. (2006). *Measuring Health: A Guide to Rating Scales and Questionnaires, Third Edition*. Oxford University Press, ISBN: 9780195165678

19. Mokhtarimousavi, S., Anderson, J. C., Azizinamini, A., and Hadi, M. (2019). Improved Support Vector Machine Models for Work Zone Crash Injury Severity Prediction and Analysis. *Transportation Research Record*. https://doi.org/10.1177/0361198119845899.

20. National Highway Traffic Safety Administration (NHTSA) (2023). Motorcycle Safety. www.nhtsa.gov/road-safety/motorcycles (Accessed May 23, 2023).

21. Rezapour, M., Farid, A., Nazneen, S., and Ksaibati, K. (2021). Using machine learning techniques for evaluation of motorcycle injury severity. *IATSS Research*, *45*(3), 277–285. https://doi.org/10.1016/j.iatssr.2020.07.004.

22. Sharma, B., Katiyar, V. K., and Kumar, K. (2016). Traffic accident prediction model using support vector machines with Gaussian kernel. *Advances in Intelligent Systems and Computing*, *437*, 1–10. https://doi.org/10.1007/978-981-10-0451-3_1.

23. World Population Review (2023). *Helmet Law States 2023*. https://worldpopulationreview.com/state-rankings/helmet-law-states (Accessed May 23, 2023).

24. Xing, E., Haleem, K., and Asce, M. (022). *Motorcycle Safety Investigation in Kentucky Using Machine and Deep Learning Techniques*. International Conference on Transportation and Development 2022, Seattle, WA

25. Yan, M., and Shen, Y. (2022). Traffic Accident Severity Prediction Based on Random Forest. *Sustainability (Switzerland)*, *14*(3). https://doi.org/10.3390/su14031729.

26. Yu, R., and Abdel-Aty, M. (2013). Utilizing support vector machine in real-time crash risk evaluation. *Accident Analysis and Prevention*, *51*, 252–259. https://doi.org/10.1016/j.aap.2012.11.027.