Motorcycle Safety Investigation in Kentucky Using Machine and Deep Learning Techniques

Eric Xing[1] and Kirolos Haleem, Ph.D., P.E., M.ASCE[2]

[1]Undergraduate Student Researcher, Gatton Academy of Mathematics and Science, Western Kentucky Univ., Bowling Green, KY. Email: eric.xing819@topper.wku.edu
[2]Assistant Professor, School of Engineering and Applied Sciences, Western Kentucky Univ., Bowling Green, KY (corresponding author). Email: kirolos60@hotmail.com

ABSTRACT

This study analyzes the factors affecting motorcycle crash severity in the state of Kentucky while applying machine learning method (i.e., random forest) and deep learning model (i.e., combined principal component-neural network model). Severe motorcycle crashes were the main severity level outcome analyzed in this study and are those crashes resulting in either serious motorcycle injury or fatality. To the authors' knowledge, these models have been very rarely implemented to analyze motorcycle crashes, especially when it comes to severe motorcycle crashes. Recent five-year motorcycle crash data (2015 to 2019) from Kentucky were used. The random forest classifier was applied to rank each feature's importance in influencing serious injury or fatal motorcycle crashes. The random forest classifier indicated that collision time, crash location, driver age, helmet use, and vehicle type colliding with the motorcycle were the key features affecting severe motorcycle crashes while yielding 91% prediction accuracy. By testing multiple numbers of principal components, 800 principal components were found to decrease overfitting while still retaining high prediction accuracy. Thus, 800 principal components were used for fitting the neural network model. The neural network showed that driver-related (i.e., age), crash-related (i.e., crash location, collision time, and manner of motorcycle collision), and roadway-related factors (i.e., roadway surface condition) could successfully predict severe motorcycle crashes with 94.2% prediction accuracy. An advanced occlusion-based interpretation of the neural network model also produced a list of features most highly correlated with the model prediction performance. The neural network model result was largely consistent with that of random forest. Nevertheless, deep learning techniques (e.g., the combined principal component-neural network model) could better predict severe motorcycle crashes with higher accuracy compared to machine learning techniques (e.g., random forest). Overall, the study results demonstrated that both machine learning (random forest) and deep learning (combined principal component-neural network model) techniques can be used successfully in identifying those key features contributing to severe motorcycle crashes.

Keywords: Motorcycle Safety, Motorcycle Crashes, Machine Learning, Deep Learning, Severe, Severity, Neural Network, Random Forest, Principal Component, Prediction Accuracy

INTRODUCTION

According to recent National Highway Traffic Safety Administration (NHTSA, 2019) crash data in 2019, while motorcycles represent only about 3% of registered motor vehicles and less than 1% of vehicle miles traveled, more than 14% of traffic fatalities in the United States involve motorcyclists. About 8.5 million motorcycles are registered in the United States. In the state of Kentucky, during the five-year period (2015 to 2019), 23% of motorcycle-related crashes resulted in either fatality or suspected serious injury (KSP, 2021). Furthermore, although motorcycle crashes in Kentucky accounted for only 1% of total crashes, they were responsible for 12% of total severe injuries in the state (including both fatal and suspected serious injuries) (KSP, 2021). The aforementioned statistics call for further motorcycle safety analysis to pinpoint those risk factors affecting severe motorcycle crashes in Kentucky.

The main objective of this study is to analyze the factors affecting motorcycle crash severity in the state of Kentucky by applying and comparing a machine learning method (the random forest technique) and deep learning technique (a combined principal component-neural network model). The previous two models were applied and compared while assessing each model's performance in effectively predicting severe motorcycle crashes in Kentucky (in order to finally recommend the best modeling approach). Note that severe motorcycle crashes are those crashes resulting in either serious motorcycle injury or fatality. In this study, recent five-year motorcycle crashes (2015 to 2019) from the Kentucky State Police collision database were used in the comparative analysis between the machine learning method (i.e., random forest technique) and deep learning technique (i.e., combined principal component-neural network model).

LITERATURE REVIEW

Motorcycle Safety & Using Machine Learning Techniques in Crash Analysis

Regarding studies that analyzed motorcycle safety, Robbins and Fotios (2020) investigated the effect of ambient light on motorcycle collisions in the UK. The authors used the odds ratio approach to isolate the effect of light and found that the risk of motorcycle collisions was much higher on roads with low speed limits, at T-shaped junctions, and at junctions with "give way" signs. Islam (2021) investigated the effect of motorcyclists' age on injury severities in single-motorcycle crashes in Florida while applying the mixed logit model. It was found that riding a motorcycle with a 10 mile per hour above speed limit increased the likelihood of fatal injury for middle age group (30 to 49 years) compared to younger age group (below 30 years). Moreover, not wearing a helmet increased the likelihood of fatal injury for older age group (50 years and above), while reduced the likelihood for middle age group. Also, analyzing the injury severity outcomes of motorcycle crashes, Tamakloe et al. (2022) used the association rules data mining technique and binary logit model to explore the factors affecting the severity of motorcycle crashes along both signalized and non-signalized intersections in Ghana. Three-year crash data (2016-2018) were used. Different factors were found to influence motorcycle severity at both intersection types. For example, the association rules technique showed that license status, daytime, and shoulder presence increased the likelihood of fatal injuries at signalized intersections, whereas inattentiveness, nighttime, shoulder absence, and young motorcyclists were found to increase motorcycle fatality risk at non-signalized intersections.

The following sections summarize studies that applied machine learning techniques in crash analysis. Abdel-Aty and Haleem (2011) utilized machine learning techniques to analyze angle crashes at unsignalized intersections in Florida. Specifically, multivariate adaptive regression splines (MARS) and negative binomial (NB) models were used and compared. The MARS model was found to outperform the NB model, and was recommended for predicting crashes at unsignalized intersections. Rezapour et al. (2021) compared the performance of various machine learning models, including random forest, support vector machines, and MARS on analyzing Wyoming motorcycle crash data (consisting of 2,484 crashes). The random forest classifier was found to have the highest prediction performance.

Harb et al. (2009) used decision trees and random forest classifier to analyze the effect of drivers, vehicles, and environmental characteristics on the presence or absence of crash avoidance maneuvers. Each manner of collision was analyzed separately. Speed limit was found to be associated with rear-end collisions' avoidance maneuvers, whereas vehicle type was correlated with head-on and angle collisions' avoidance maneuvers. Chang et al. (2019) employed the classification and regression trees (CART) to analyze the effect of human illegal human behaviors on the occurrence of 4,587 motorcycle crashes involving serious injury or fatality in Hunan, China. The authors found that helmet use was one of the key factors affecting a crash severity outcome. Recently, Rezapour et al. (2020) also used CART to identify the key factors in predicting severe and fatal outcomes in 1,360 at-fault motorcycle crashes. Speed limit, driver age, highway functional class, and speed compliance were the key factors affecting motorcycle injury severity.

Literature Review on Using Deep Learning and Text Mining in Crash Analysis

While applying a deep learning technique, named *DeepScooter*, for classifying motorcycle crash injury severity, Das et al. (2018) could achieve a classification prediction accuracy of 94%. Arteaga et al. (2020) investigated the description of heavy vehicle crashes in Queensland, Australia as a potential source of fatal crash-causing factors. The authors used the Global Cross-Validation Local Interpretable Model-Agnostic Explanations (GCV-LIME) text mining or deep learning approach that combines the Global Cross-Validation model with the LIME model architecture. The proposed model could identify head-on collisions, side of collision, motorcycles, cabs, and pedestrian as highly-correlated factors with fatal crashes.

Most recently, Kwayu et al. (2021) analyzed fatal crashes in Michigan using structural topic modeling (STM) and network topology approaches, which are deep learning-based models. Data were sorted into themes related to pre-crash events, crash locations, and involved parties. These themes have included angle and type of collision, crash near stop signs, crash crossing the centerline, and vehicle's inability to stop. The Eigenvector centrality was used to find consistent themes, and a high crash prediction performance was finally detected from the generated themes.

Study Contribution to the Literature

The aforementioned review of literature has shown that machine and deep learning approaches have been widely applied and used in general safety and crash investigations. To the authors' knowledge, the application of such advanced techniques in motorcycle crash analysis has been relatively limited, especially when it comes to severe motorcycle crash analysis. This study takes the initiative and analyzes the factors affecting severe motorcycle crash outcomes in the state of Kentucky by applying and comparing machine and deep learning techniques. Specifically, the study adds to the current body of literature by applying advanced approaches to help improve severe motorcycle crash prediction. Note that the random forest classifier was used to represent the machine learning technique, whereas the combined principal component-neural network modeling approach was used to represent the deep learning technique. The prediction performance from each approach was compared and assessed. Furthermore, the key factors affecting severe motorcycle crashes were identified using the "Gini node impurity index" (as part of the random forest technique) and the "prediction accuracy after feature removal" (as part of the neural network model). Recent five-year motorcycle crash data (2015 to 2019) in Kentucky were used in the comparative analysis in this study.

DATA COLLECTION AND PROCESSING

Crash Data Acquisition

Five-year statewide motorcycle crash data (2015 to 2019) were extracted from the Kentucky State Police (KSP, 2021) database. Crashes were screened out to only include motorcycle-related crashes. Crash data from 2020 were omitted due to the potential confounding effect of the COVID-19 pandemic on the data. After doing some data cleaning processes, the final number of motorcycle crashes was 5,005. This sample size is considered greater than (or comparable to) the crash sample sizes in previous studies that applied machine and deep learning techniques (see for examples, *1,360-related crashes* by *Rezapour et al. 2020*, *2,484-related crashes* by *Rezapour et al. 2021*, and *4,587-related crashes* by *Chang et al. 2019*). Each crash included related information such as environmental conditions, driver, and geolocation. Table 1 shows summary descriptive statistics of the motorcycle crash data used for key variables investigated in the study. For illustration purposes, Figure 1 displays a histogram for the distribution of "total and severe" motorcycle crashes by different age groups. Note that "very young age" included "15 to 19 years old", "young age" included "20 to 24 years old", "middle age" included "25 to 64 years old", "old age" included "65 to 79 years old", and "very old age" included "80+ years old". Interestingly, from Figure 1, "middle age group" had the highest contribution of serious (or severe) motorcycle crashes.

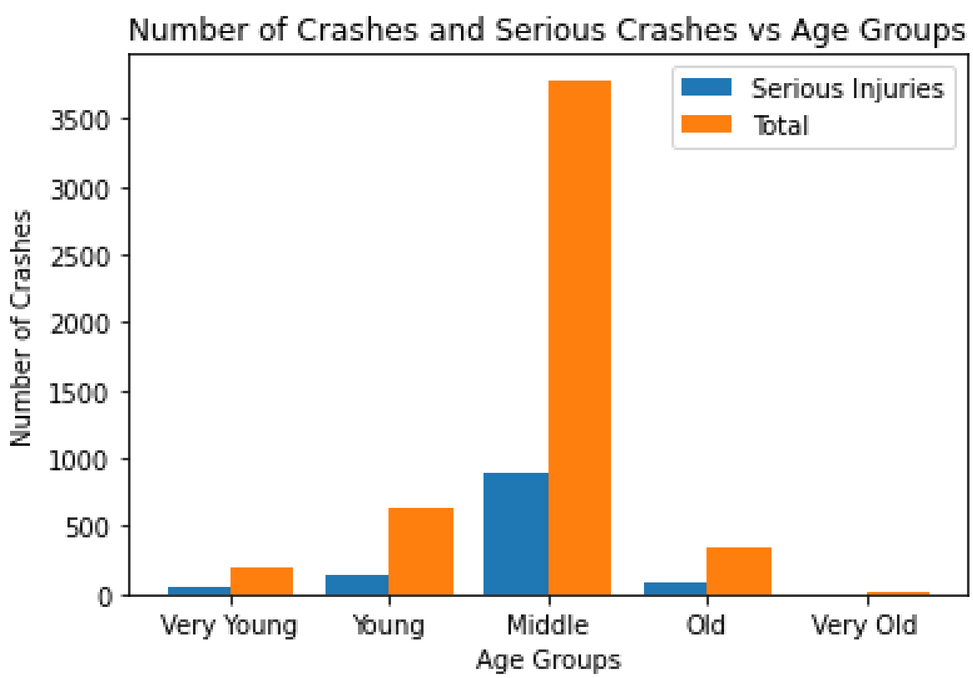

Figure 1. Distribution of motorcycle crashes by different age groups.

Table 1. Summary Descriptive Statistics of Motorcycle Crashes in Kentucky.

Variable	Variable Levels	Motorcycle Crash Frequency
Crash Injury Severity	Severe (K+A)	1,216
	Non-Severe (B+C+O)	3,789
Gender	Male	4,722
	Female	283
Helmet Use	Helmet Used	2,495
	Helmet Not Used	2,510
Number of Motorcycle Deaths	0	4,711
	1	80
	≥ 2	214
Number of Serious Motorcycle Injuries	0	4,083
	1	599
	2	281
	≥ 4	42
Weather Condition	Clear	4,148
	Cloudy	659
	Raining	148
	Fog	35
	Other	15
Roadway Surface Condition	Dry	4,672
	Wet	305
	Other	28
Hit-and-Run Crash	No	4,821
	Yes	184
Roadway Type	Straight & Level	2,401
	Curve & Level	844
	Curve & Grade	703
	Straight & Grade	655
	Straight & Hillcrest	235
	Curve & Hillcrest	167
Lighting Condition	Daylight	3,774
	Dark (Lighted)	445
	Dark (Not Lighted)	501
	Dusk	156
	Dawn	82
	Unknown	47

Data Cleaning

Data were initially present in a CSV file with multiple sheets, each sheet containing all available data about one specific aspect of a specific crash, such as persons involved or environmental data. Information from different sheets were merged using the unique matching crash ID. Person data with the motorcycle driver tag were given higher precedence than nonmotorcycle driver persons. For data without associated motorcycle driver data, the first motorcycle occupant was chosen. Note that severe motorcycle crashes were the main severity level outcome analyzed in this study and they are defined as those crashes resulting in either serious injury or fatality.

Each feature (or explanatory variable) was examined to determine its missing information proportion. If the proportion was over 0.25, that feature was excluded for the analysis. A correlation matrix was then constructed for the remaining features, which was used to identify pairs of variables with high correlations. One value out of each pair was removed in order to reduce data size while maintaining much of the original data variance. For features with sparingly missing values, missing values were replaced will null values. For most categorical variables, one-hot encoding was applied, converting them to a sequence of dummy binary variables, indicating the presence or absence of a specific case of the categorical variable class. The data were then split into a training and testing sets, in a manner that maintained approximately equal proportions of severe crashes in each subset of the data. In this study, the training/testing data split was set at 80/20. Note that before passing the training data into learning algorithms, the data were normalized with Scikit-Learn's Standard Scaler to remove the effect of certain inherently large feature values. After the mean and standard deviation of each feature were calculated, the following transformation was applied to each feature value (note that the normalization produced values that are normally distributed between 0 and 1, which were then fed to the neural network model, as will be shown later):

$$standardized\ value = \frac{origianal\ value - feature\ mean}{feature\ standard\ deviation} \tag{1}$$

METHODOLOGY

Random Forest Classifier

A random forest classifier is a machine learning technique, and is defined as an ensemble of uncorrelated decision trees that produce an averaged output. Scikit-Learn's implementation of the random forest classifier in Python was used in this study. The random forest classifier generally has a much higher performance than a single decision tree, as it consists of multiple relatively uncorrelated decision trees.

Two main features are relatively unique to the random forest. The first is the use of bootstrap aggregation. Each decision tree is trained on a separate subset of the original input data. This subset is a random sample with a replacement of the original input data. This makes each decision tree very sensitive to changes in the input data and allows the averaged results of all decision trees in the forest to be more accurate. Bootstrap aggregation also addresses the fundamental weakness of the decision tree, high variance. Bootstrap aggregation, combined with a large number of decision trees results in generalization of data, can reduce overfitting and lower variance.

Additionally, each decision tree operates on a randomly selected subset of the original data's features. The number of decision trees is determined via comparison of "Out-of-Bag Error" (OOB error). OOB error is defined as the mean prediction error on a data entry *s* over all trees that did not include *s* in their bootstrap sample. It measures the power of a decision tree to generalize the training data to testing data. The optimal number of trees is the lowest number that produces the most optimal results. The importance of each feature in a dataset can be relatively determined via the calculation of Gini impurity. A higher Gini node purity correlates to a high feature importance. Feature importance was calculated in the form of the mean decrease in "Gini node impurity" or the contribution of the feature in increasing the prediction performance.

Principal Component Analysis & Neural Network Model

The principal component analysis is a dimensionality reduction technique for reducing the size of data. Additionally, principal component analysis reduces the degree of meaningless noise in the data; hence, it reduces the degree of data overfitting by the neural network model. Essentially, principal component analysis retains the maximal degree of data variance after a transformation to a lower dimension. This is done by re-dimensionalizing the data with a number of unit vectors called principal components, where the number of principal components is less than the original number of features. Each principal component is defined as the unit vector that most closely follows the least squared regression line while being orthogonal to all previous principal components. Once a certain number of principal components are constructed, the data were dimensionalized. Specifically, the Scikit-Learn implementation was used as part of the principal component analysis in Python. Scikit-Learn is a free software machine learning library in the Python programming language.

After dimensionalizing severe motorcycle crashes in this study using the principal component analysis, the neural network model was implemented afterwards. Neural network is a machine learning technique and neural networks are a collection of nodes grouped in layers, where each node receives data from the previous layer and transmits a value to the subsequent layer. The neural network implemented in this study is a fully-connected neural network, where each node receives the output from all nodes in the previous layer and transmits data to all nodes in the subsequent layer. The neural network was implemented in Python using the PyTorch library. Note that the research team did not use PyTorch's models with a predefined architecture, but instead defined a custom-sized neural network consisting of 10 layers of 2048 neurons.

Neural networks generally consist of an input layer, hidden layers, and an output layer. Input data are passed into the input layer, where each node takes the value of a specific feature. The hidden layers perform computations and transformations on the input data that transform it into the output. The output layer provides prediction probabilities for each prediction class, where the highest probability is the outputted guess of the network. Figure 2 demonstrates the complete interconnectivity and layered structure of fully-connected neural networks.

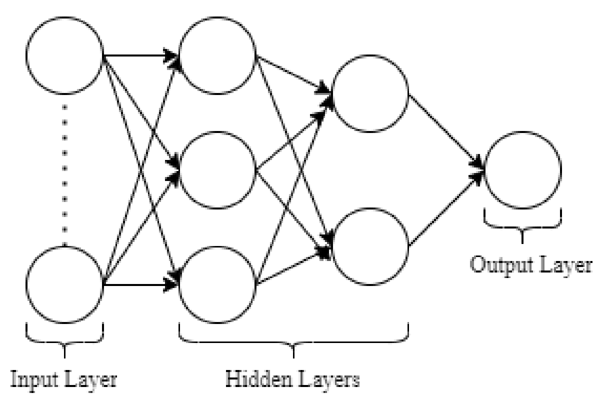

Figure 2. A generic fully-connected neural network architecture.

At each node, a polynomial-based mathematical transformation is applied to the received input to produce the node output. This high degree polynomial is written in terms of the outputs of the nodes in the previous layer. This polynomial produces some value, *x*, which is further transformed by the activation function. In this case, the activation function was defined to be the Rectified Linear Unit (or ReLU) function, which is expressed as:

$$ReLU(x) = \max(0, x) \tag{2}$$

Note that this transformation does not change the value of a positive input; however, transforms any negative input to 0. In this study, a metric of inaccuracy, called the loss function, was defined. This function aims to maximize the neural network's performance and accuracy. The loss function used in this paper is Cross-Entropy loss, which is defined as:

$$loss = -\sum_{i=1}^{n} y_i * \log(\hat{y}_i) \tag{3}$$

where: n is the number of classes and y_i is the predicted value for class i.

In this study, the value of the loss was minimized through a process called gradient descent, where the parameters of each node were updated as the loss was propagated backwards from the output layer for each testing datum. These updates were made in order to decrease the loss function, in order to converge into a local or global minimum of the loss function. Afterwards, the gradient descent was terminated, since the loss function has already reached its minimum. Since the neural network model can achieve a baseline performance of 76.3% via a constant negative output, regardless of the input data, the loss function was weighted as the default loss multiplied by the frequency of the class in the data. So, the updated (or weighted) loss function is as follows:

$$wt.d\ loss = -\sum_{i=1}^{n} (y_i * \log(\hat{y}_i) * (1 - pred) * 0.763) + (y_i * \log(\hat{y}_i) * pred * 0.237) \tag{4}$$

where: 0.763 and 0.237 are the frequencies of the positive and negative class, respectively. The process of gradient descent was repeated for each testing datum in the testing portion of the dataset. The neural network models were trained for 10 epochs, where each epoch consisted of performing gradient descent for all data in the training portion of the dataset and determining the model performance (i.e., model validation) by calculating the prediction accuracy on the validation portion of the dataset.

Note that in order to analyze the importance of specific features in the neural network model prediction, an occlusion-based approach was employed. This approach was inspired by the window occlusion approach described by Fergus (2013). A specific feature of interest was removed before principal component analysis, and the re-dimensionalized data were used to train the model ten times in order to obtain the overall average accuracy. The reduction in accuracy compared to the baseline is approximately proportional to the feature's importance.

RESULTS AND DISCUSSION

Random Forest Technique

With an 80/20 split (i.e., training/testing data split), the random forest model could predict severe motorcycle crashes with an accuracy of 91%. This performance is consistent (yet still higher) than previous literature (e.g., Rezapour et al. 2021, who got 86% prediction accuracy while analyzing Wyoming's motorcycle crashes using the random forest technique).

Afterwards, feature ranking was employed to identify the most critical features affecting motorcycle crash injury severity (also used in previous studies, e.g., Abdel-Aty and Haleem 2011). Figure 3 displays the top ten features ranked by importance vs. relative importance (as described by mean decrease in Gini node impurity). As shown, the results indicated that collision time, crash location, driver age, helmet use, and number of vehicles involved in the motorcycle crash were the five most important factors to predict severe motorcycle crashes. Noticeably, other studies, e.g., Rezapour et al. (2021), found that driver age was one of the most significant factors affecting motorcycle crashes in Wyoming. The other five important factors (from six to ten), in order were: motorcycle crash with pedestrians at non-intersections, motorcycle collision along straight and hillcrest roads, left-turning motorcycle-related collision, backing motorcycle-related collision, and dawn lighting condition.

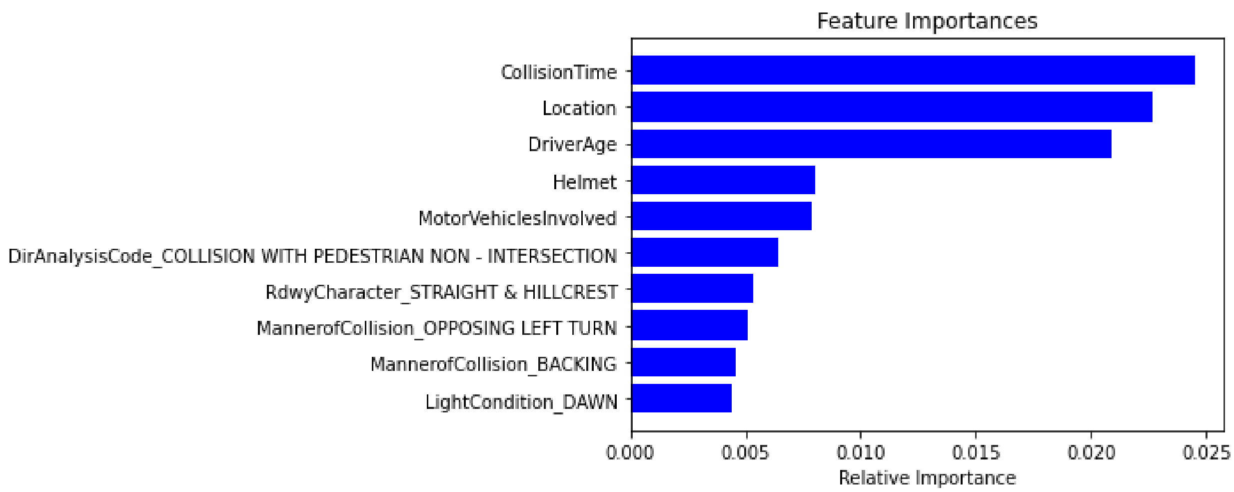

Figure 3. Relative importance of ten most important features affecting severe motorcycle crashes (ranked using mean Gini node impurity decrease).

Principal Component and Neural Network Model

In order to ensure that an acceptable amount of data was retained during dimensionalizing with principal components, the cumulative explained variance, a metric of retained information, was calculated for each number of principal components up to 1000 principal components, at which point the number of principal components exceeded the original number of features. Figure 4 demonstrates that with around 800 principal components, a high degree of cumulative explained variance was retained, while input size was reduced. By testing multiple numbers of principal components, it was determined that 800 principal components decreased the degree of overfitting while still retaining high performance. Therefore, 800 principal components were used for re-dimensionalizing the data and fitting the neural network model, as shown next.

After running the neural network model using Python (while having severe motorcycle crash level as the response variable), the visual output of the model is represented in Figure 5. As demonstrated by Figure 5, the neural network model achieved a 94.2% testing prediction accuracy after 10 epochs, which is higher than that from the random forest classifier. The 94.2%

prediction accuracy is also slightly higher than the prediction accuracy reported by Das et al. (2018) who used a different deep learning technique, the *DeepScooter*. Note that a higher accuracy (around 98%) were gained when using more than 10 epochs; however, overfitting was found to occur. For this, the neural network model was not trained after 10 epochs, and 10 epochs were set as the threshold.

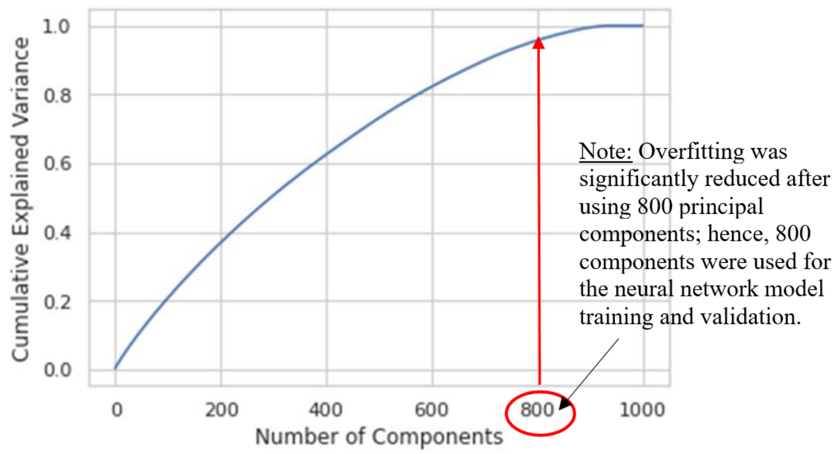

Figure 4. Cumulative explained variance plot vs. number of principal components.

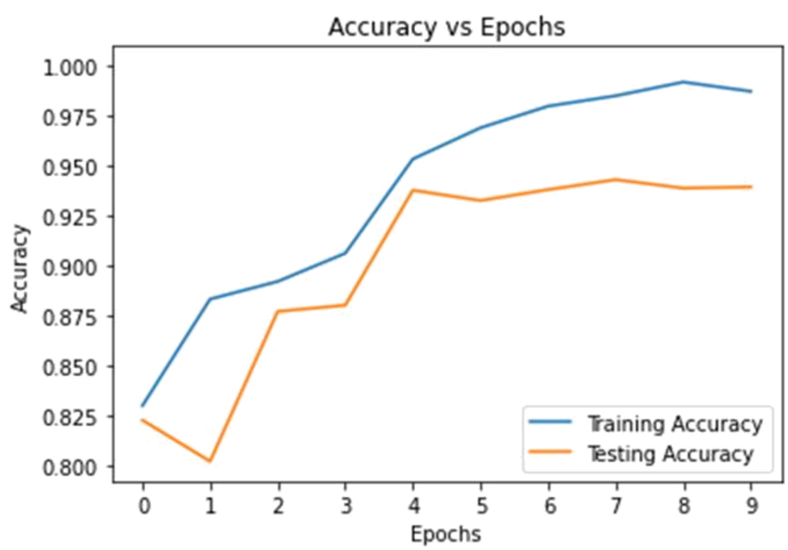

Figure 5. Training and validation accuracy vs. number of epochs.

Table 2 shows the prediction accuracies of the different trained neural network models after removal of specific features or variables (using the occlusion-based interpretation technique), the percent reduction in prediction accuracy (from the 94.2% baseline accuracy), and the order of feature importance. To better understand this table, those features removed from the neural network model and finally producing the least prediction accuracy (or, in other words, the highest percent reduction in prediction accuracy) were deemed as the most important features in predicting severe motorcycle crashes.

As shown in Table 2, the top five most important features, in order, were: driver age, crash location, collision time, manner of motorcycle collision, and roadway surface condition. The next four most important features were: number of vehicles in motorcycle crash, helmet use, driving under influence of alcohol, and lighting condition. This result is largely consistent with the ranking produced by the random forest classifier (specifically, diver age, crash location, and collision time).

Table 2. Severe Motorcycle Crash Prediction Accuracy after Using Occlusion-Based Interpretation as part of the Neural Network Model (with Specific Feature Removal).

Feature/Variable	Prediction Accuracy after Feature Removal (%)	% Reduction from Baseline Prediction Accuracy of 94.2%	Order of Feature Importance
Driver Age	89.8%	4.4%	1
Crash Location	90.0%	4.2%	2
Collision Time	90.4%	3.8%	3
Manner of Motorcycle Collision	91.4%	2.8%	4
Roadway Surface Condition	91.6%	2.6%	5
Number of Vehicles Involved in Motorcycle Crash	91.9%	2.3%	6
Helmet Use	92.0%	2.2%	7
Driving under Influence of Alcohol	92.0%	2.2%	7
Lighting Condition	92.1%	2.1%	9

CONCLUSIONS AND RECOMMENDATIONS

This study investigated severe motorcycle crashes in Kentucky while applying machine learning method (i.e., the random forest classifier) and deep learning model (i.e., the combined principal component-neural network model). To the authors' knowledge, these models have been rarely implemented to analyze motorcycle crashes, especially when it comes to severe motorcycle crashes. Five-year (2015 to 2019) statewide motorcycle crashes in Kentucky were used in the analysis. The random forest results indicated that collision time, crash location, driver age, helmet use, and number of vehicles involved in the motorcycle crash were the five most important factors to predict severe motorcycle crashes. Furthermore, the random forest could achieve a high severe motorcycle crash accuracy of 91%.

After using the principal component method and while testing multiple numbers of principal components, it was found that 800 principal components could reduce the degree of overfitting while still retaining high performance. For this, 800 principal components were used for fitting the neural network model. The neural network demonstrated that driver-related (i.e., age), crash-related (i.e., crash location, collision time, and manner of motorcycle collision), and roadway-related factors (i.e., roadway surface condition) could successfully predict severe motorcycle crashes. Note that this ranking was largely consistent with that produced by the random forest classifier (specifically, diver age, crash location, and collision time). In addition, the neural network model achieved a testing prediction accuracy of 94.2%, which is higher than that from the random forest classifier.

Overall, the neural network model result was largely consistent with that of random forest classifier, which attests to the success of using machine and deep learning techniques in modeling and predicting severe motorcycle crashes. Nevertheless, if safety researchers and practitioners need to achieve a higher prediction accuracy, the use of deep learning techniques (e.g., the combined principal component-neural network model) could better predict severe motorcycle crashes compared to machine learning techniques (e.g., random forest) while attaining a higher prediction accuracy. This was mainly since the fitness of deep learning techniques on manipulating severe motorcycle crashes seemed inherently stochastic, which resulted in a better model performance.

It is recommended to apply deep learning methods when analyzing severe motorcycle crashes and identifying the significant variables affecting these crashes to better monitor those variables and improve motorcyclists' safety. From the identified significant variables affecting severe motorcycle crashes in this study, it is suggested to restrict motorcycle traffic during certain times of day at those high motorcycle crash spots (e.g., at non-peak periods), separate motorcycle traffic from regular vehicular traffic (by adding motorcycle-dedicated lanes whenever possible), and have stricter enforcement of motorcycle helmet use. Future research could expand upon this study by exploring other deep learning models (e.g., support vector machines) in analyzing severe motorcycle crashes and then comparing the results with the neural network model. Moreover, other model interpretability algorithms could be used in order to perform more precise feature ranking on the neural network deep learning model.

One limitation of this study is the relatively small motorcycle crash sample size used. Nevertheless, as previously noted, the current sample size is still deemed reasonable (compared to previous studies). It will be still interesting to apply the used machine and deep learning models in this study to a larger crash sample size (e.g., with at least 10,000 records) and then conduct a sensitivity-type analysis to compare the finally-obtained results with the current study findings.

REFERENCES

Abdel-Aty, M., and Haleem, K. (2011). "Analyzing Angle Crashes at Unsignalized Intersections Using Machine Learning Techniques." *Accident Analysis & Prevention*, 43(1), 361-470.

Arteaga, C., Paz, A., and Park, J. W. (2020). "Injury Severity on Traffic Crashes: A Text Mining with an Interpretable Machine-Learning Approach." *Safety Science*, 132, 104988.

Chang, F., Xu, P., Zhou, H., Lee, J., and Huang, H. (2019). "Identifying Motorcycle High-Risk Traffic Scenarios through Interactive Analysis of Driver Behavior and Traffic Characteristics." *Transportation Research Part F: Traffic Psychology and Behaviour*, 62, 844-854.

Das, S., Dutta, A., Dixon, K., Minjares-Kyle, L., and Gillette, G. (2018). "Using Deep Learning in Severity Analysis of at-Fault Motorcycle Rider Crashes." *Transportation Research Record*, 2672(34).

Fergus, R. (2013). Visualizing and Understanding Convolutional Networks (last revised Nov. 28, 2013). <https://arxiv.org/abs/1311.2901>(January 19, 2022).

Harb, R., Yan, X., Radwan, E., and Su, X. (2009). "Exploring Precrash Maneuvers Using Classification Trees and Random Forests." *Accident Analysis & Prevention*, 41(1), 98-107.

Islam, M. (2021). "The Effect of Motorcyclists' Age on Injury Severities in Single-Motorcycle Crashes with Unobserved Heterogeneity." *Journal of Safety Research*, 77, 125-138.

Kentucky State Police Web Portal. <https://crashinformationky.org>(August 22, 2021).

Kwayu, K., Kwigizile, V., Lee, K., and Oh, J.-S. (2021). "Discovering Latent Themes in Traffic Fatal Crash Narratives Using Text Mining Analytics and Network Topology." *Accident Analysis & Prevention*, 150, 105899.

NHTSA (National Highway Traffic Safety Administration). Traffic Safety Facts 2019. <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813141>(August 22, 2021).

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.

Rezapour, M., Molan, A., and Ksaibati, K. (2020). "Analyzing Injury Severity of Motorcycle at-Fault Crashes Using Machine Learning Techniques, Decision Tree and Logistic Regression Models." *International Journal of Transportation Science and Technology*, 9(2), 89-99.

Robbins, C., and Fotios, S. (2020). "Motorcycle Safety after-Dark: The Factors Associated with Greater Risk of Road-Traffic Collisions." *Accident Analysis & Prevention*, 146, 105731.

Tamakloe, R., Das, S., Aidoo, E., and Park, D. (2022). "Factors Affecting Motorcycle Crash Casualty Severity at Signalized and Non-Signalized Intersections in Ghana: Insights from a Data Mining and Binary Logit Regression Approach." *Accident Analysis & Prevention*, 165, 106517.