2021 IEEE Global Conference on Artificial Intelligence and Internet of Things (GCAIoT) | 978-1-6654-3841-4/21/$31.00 ©2021 IEEE | DOI: 10.1109/GCAIoT53516.2021.9693055

Road Accident Severity Prediction — A Comparative Analysis of Machine Learning Algorithms

Sumbal Malik[a,b,*], Hesham El Sayed[a,b], Manzoor Ahmed Khan[a,b], Muhammad Jalal Khan[a]
[a]College of Information Technology, UAE University, Al Ain, Abu Dhabi, United Arab Emirates
[b]Emirates Center for Mobility Research (ECMR), UAE University
{201990107,helsayed,manzoor-khan,201990067}@uaeu.ac.ae

***Abstract*—Crash severity prediction models enable various agencies to predict the severity of a crash to gain insights into the factors that affect or are associated with crash severity. One of the potential ways to predict the crash severity is to leverage machine learning (ML) algorithms. With the help of accident data, ML algorithms find hidden patterns to predict whether the severity of the crash is fatal, serious, or slight. In this research, we develop a prediction framework and implemented six different machine learning algorithms, namely: Naïve Bayes, Logistic Regression, Decision Tree, Random Forest, Bagging, and AdaBoost to predict the severity of the crash. Experimental results procured for the crash dataset published by the UK shows that Random Forest, Decision Tree, and Bagging significantly outperformed other algorithms in terms of all performance metrics. Furthermore, we analyze the huge; traffic data and extract insightful crash patterns to figure out the significant factors that have a clear effect on road accidents and provide beneficial suggestions regarding this issue. We strongly believe that the proposed prediction framework and the extracted pattern analysis would be helpful in improving the traffic safety system and assist the road authorities to establish proactive strategies to prevent traffic accidents.**

***Index Terms*—crash severity; machine learning, prediction, random forest, logistic regression**

I. Introduction

Road accidents have become one the biggest problems in the world [1]. It brings immense economic, physical, and emotional suffering to society. The World Health Organization (WHO) states that more or less 1.35 million people die yearly and more than 50 million are injured globally due to road accidents [2]. Furthermore, the statistics show that road accidents are the leading cause of death for children and young people aged between; 5 - 29 years [3]. This alarming situation pushes us to research and improve the processes aiming to help traffic authorities in enhancing road safety and evaluating the factors that contribute to the accidents.

Crash severity prediction is undoubtedly a fundamental aspect of a crash event. Therefore, providing accurate and timely predictions of the crash severity helps in reducing, the response time for emergency units, enhance road safety, and reduce traffic congestion. Recently, a substantial amount of research has been done in predicting the crash severity leveraging various solution approaches such as descriptive and inferential statistical analysis. However, machine learning (ML) algorithms have appeared as one of the potential approaches for modeling crash severity due to their excellent outcomes. Furthermore, ML algorithms provide tremendous flexibility, are non-parametric tools requiring little assumptions about crash severity data, and can greatly deal with handling missing values, noises, and outliers.

This research aims to implement a crash prediction framework, leveraging machine learning algorithms to predict whether the crash is fatal, serious, or slight. The objectives of this research are manifold: i) to implement the crash severity prediction framework; ii) to compare several machine learning algorithms to determine the best algorithm with the highest performance metrics; iii) to mine and analyze the huge traffic data to extract the useful hidden crash patterns and the factors contributing to the crashes.

This paper is structure into six sections. Section II presents an overview of the related work. Section III details the proposed framework and methodology. Section IV analyzes the experimental results and discusses the effectiveness of our crash severity prediction framework. Section V discusses the insightful crash patterns. Finally, Section VI elucidates the conclusion of the study.

II. Related Work

In this section, the state-of-the-art literature related to crash severity prediction is discussed. The research study conducted by Najada et al., [4] used Hong Kong's transportation dataset to predict the actual causes of the accidents. They implemented and compared the performance of several classification algorithms in WEKA. Their experimental results showed that Random Forest surpassed the Naïve Bayes and PART algorithm. Similarly, in another study, Chong et al., [5] built the model using ML algorithms to classify the accident injury severity into five categories. They used Artificial Neural Network (ANN), Support Vector Machine (SVM), and Decision Tree to develop the model. The results of their study highlighted that speeding over the lawful limit is the main reason for fatal injury. Moreover, another interesting research study carried out by the authors [6], helped the community in understanding how accident prediction is done using big data mining and data analysis. In addition, the authors also highlighted how to use data sampling to reconstruct the dataset followed by using prepossessing techniques to make data complete and reliable.

Authorized licensed use limited to: DUBLIN CITY UNIVERSITY. Downloaded on February 03,2024 at 23:13:54 UTC from IEEE Xplore. Restrictions apply.

The research study conducted by Elfar et al., [7], used three machine learning algorithms: logistic regression, random forest, and neural networks to predict traffic accidents and congestion. Furthermore, they also proposed two predictive models for training the data. The proposed models achieved more accuracy over the other three classification models used in their study. The authors' contributions further guided how their proposed models can be used in various vehicular applications to improve road safety by warning drivers of upcoming traffic slowdowns. Iranitalab [8], experimented that ML algorithms such as linear regression, Naive Bayes, and Random Forrest are effectively used to analyze huge datasets to predict road accidents. They designed and implemented a framework leveraging the above-mentioned ML algorithms along with a versatile architecture namely Lambda Architecture to execute their proposed framework faster and accurately. The majority of the literature review used either simulated environments or small datasets to evaluate their findings.

In another research study, the researchers [9] suggested that neural networks are effectively being used to identify the behavioral changes of the drivers to avoid major accidents. The authors used two deep learning models, the Long Short-Term Memory (LSTM) and Recurrent Neural Network (RNN). They trained the model by using individual drivers' patterns including acceleration, deceleration, and speed. The results showed that the model effectively classified safe and unsafe driving. In recent research [10], the authors studied the challenges associated with the transportation system. Moreover, they designed the real-time transportation data mining framework by utilizing the transportation dataset of the UK from 2005 – 2012. Due to the limited resources, they only used 5486 instances for both training and validation phases. Despite, this research provided a good analysis of traffic crashes, however, the incomplete dataset could not achieve good results.

By reviewing the literature thoroughly, we highlight manifold research gaps in the existing crash severity prediction systems: i) most of the previous research studies predicted the accident associated with one or two factors only which are not sufficient for a real situation [11]; ii) a significant number of studies does not deal with the class imbalance problem; iii) unobserved heterogeneity; iv) most of the studies solely rely on a single accuracy measure to evaluate the performance of the algorithm. Therefore, this research aims to fill all the aforementioned research gaps.

III. Methodology

This research aims to build a predictive model to predict crash severity. The proposed framework illustrated in Fig. 1 consists of the following steps; i) the traffic accident data is downloaded, provided by the Department of Transportation UK; ii) the data is pre-processed and cleaned to build models; iii) the cleaned dataset is then, divided into training and testing sets; iv) the predictive models are built for six machine learning algorithms to predict crash severity; v) testing data is given to the model to test the model performance; vi) at this step, the crash severity is predicted by the model; vii) the results of all the algorithms are evaluated and compared.

A. Dataset

In this research, the road accident dataset published by the Department of Transport UK [12] is used to predict crash severity. The dataset consists of 1,22,636 instances and 32 attributes. Some of the main attributes and their description are discussed in Table I.

TABLE I
Brief Description of Road Accident Dataset

Sr	Attribute	Description
1	Index	Index of the traffic accident.
2	Longitude	Longitude of the location of an accident scene.
3	Latitude	Latitude of the location of an accident scene.
4	Accident Severity	Severity of the accident: fatal, serious or slight.
5	Vehicles	Number of vehicles involved in the accident.
6	Casualties	The number of person injured in the accident.
7	Date	Date of the accident.
8	Time	Timestamp of the accident.
9	Week Day	Day of the week that accident occurred.
10	Road Type	The type of road where the accident occurred.
11	Speed Limit	Speed limit of road where accident occurred.
12	Weather	Weather condition at the time of the accident.
13	Light Conditions	Light conditions at the time of the accident.
14	Rular /Urban Area	Area where the accident occurred.

B. Preprocessing

Data preprocessing and cleaning are the most important part of machine learning modeling. Without preprocessing, the raw data cannot be transformed into a useful and efficient format for creating a machine learning model.

The road accident dataset is available in CSV format. After downloading the dataset, we remove the irrelevant and duplicate attributes. After then, we perform the attribute value transformation from string to numeric values. Lastly, we normalize the data and convert it into the ARFF [13] format. Normalization is a technique to pre-process the data by bringing all the attributes under the common scale which is under the common minimum, maximum and medium values without distorting the differences in the ranges of values. To normalize the road accident dataset we use the Z-Score Normalization technique [14]. This technique uses the mean and standard deviation of each feature of training data to normalize each input feature vector. The Eq. (1), formalizes the Z-Score normalization where Z represents the normalized attribute value, x_i original attribute value, μ mean, and σ is the standard deviation.

$$Z = \frac{x_i - \mu}{\sigma} \tag{1}$$

Since the crash dataset is imbalanced, therefore, training an ML model with this imbalanced dataset often causes the model to develop a certain biasness towards the majority class. To avoid this biasness, the Synthetic Minority Oversampling Technique (SMOTE) is used to balance the data properly. In SMOTE, the instances of a minority class of the dataset are increased by creating new synthetic instances rather than by

Authorized licensed use limited to: DUBLIN CITY UNIVERSITY. Downloaded on February 03,2024 at 23:13:54 UTC from IEEE Xplore. Restrictions apply.

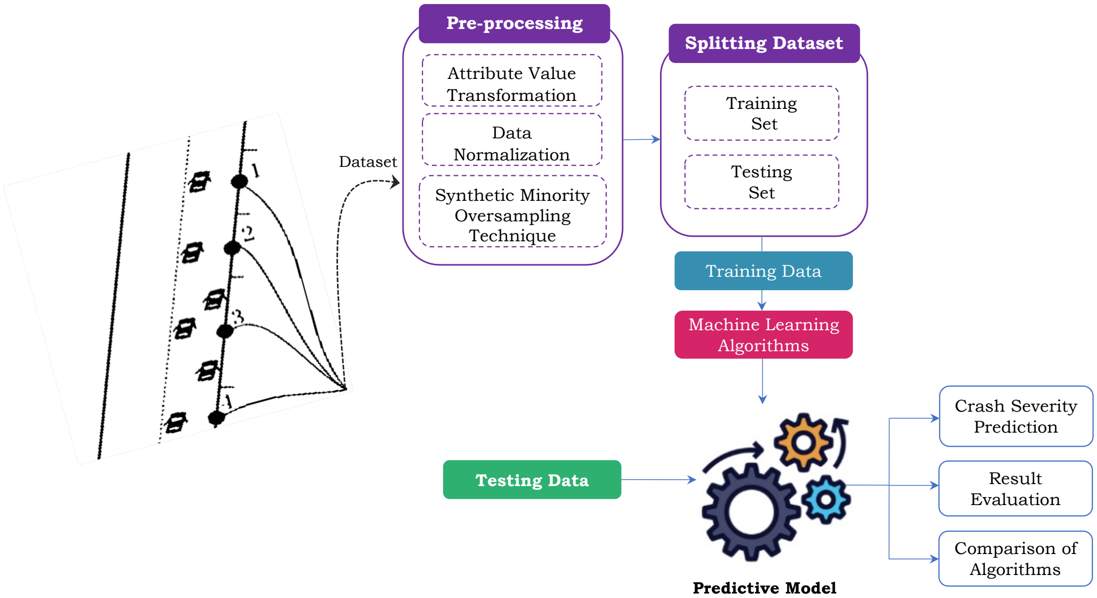

Fig. 1. Crash Severity Prediction Framework.

replication; hence it avoids the overfitting problem in ML algorithms. These instances are crafted by considering two parameters i) oversampling rate; ii) the number of nearest neighbors K. In this study, we applied the SMOTE with the following parameters: *nearestNeighbours = 5*, and *randomSeed value = 1*. For each minority instance, the algorithm takes $K = 5$ nearest neighbors (closest in distance) to the minority class (Euclidean distance). The SMOTE generates new synthetic instances ($\text{Instances}_{\text{new}}$) by taking the difference between variables of the minority sample and its nearest neighbors. Finally, the distance is then multiplied by a random number δ (between 0 and 1) and added to the variable's value of the minority sample.

C. Machine Learning Algorithms for Prediction

After pre-processing the data, we split the data into 70% training and 30% testing sets. The training data is given to ML algorithms to train the model. We build the prediction models, and the crash severity class attribute is used as the target variable. Six ML algorithms: Naïve Bayes [15], Random Forest [16], Bagging, Decision Tree [17] [18], Logistic Regression [19], and AdaBoost are trained, to predict crash severity. Table II. presents the selected parameters used to train the algorithms. After training the classifiers, the testing data is given to the model to predict crash severity and compare the results of the algorithms.

IV. Experimental Results and Discussion

In this section, the experimental results are discussed in detail. We conduct the experiments using Windows 10 operating system and the Waikato Environment for Knowledge Analysis (WEKA) 3.8.1 version. We used five performance evaluation metrics to evaluate the results of the crash severity framework. The description and formulation of these metrics are shown

TABLE II
Selected parameters for machine learning algorithms.

Algorithm	Parameters
Decision Tree	Batch size = 100; Confidence Factor = 0.25; Seed = 1; Laplace = False
Bagging	Classifier = RepTree; Iterations = 10; Seed = 1
Random Forest	Batch size = 100; Iterations = 100; Seed = 1
AdaBoost	Classifier = Decision Stump; Iterations = 10; Seed = 1; Resampling = False; Weight Threshold = 100

in Table III. These metrics are evaluated using a confusion matrix where True Positive (TP) is the crash belonging to class 1 correctly classified into the same class and True Negative (TN) is the instances belonging to class 0 correctly classified into the same class. False Negative (FN) is the instances belonging to class 1 incorrectly classified into class 0 whereas; the False Positive (FP) is the instances; belonging to class 0 incorrectly classified into class 1.

The experimental results of the crash severity prediction framework are shown in Table IV. It is evident from Table IV and Fig 2 that naive Bayes, random forest, and AdaBoost take less computational time 0.59, 4.74, and 7.05s respectively to predict the crash severity. Besides, the results also indicate that the random forest, decision tree, bagging, and AdaBoost outperformed the other algorithms in terms of all performance metrics. The results show that these machine learning algorithms can predict more accurately whether the severity of the crash is fatal, serious, or slight.

The highest accuracy obtained by implementing the algorithms listed in Table IV shows that algorithms of the tree category surpassed the probabilistic category. The random forest attained 98.80% accuracy followed by 98.4% precision, 99% recall, and 98.6% f-measure. The logic behind the random

Authorized licensed use limited to: DUBLIN CITY UNIVERSITY. Downloaded on February 03,2024 at 23:13:54 UTC from IEEE Xplore. Restrictions apply.

TABLE III
PERFORMANCE METRICS DESCRIPTION AND FORMULATION

Metrics	Description	Formula
Time	Computes the time algorithm takes to run	-
Accuracy	Computes the accuracy of the algorithm which is the rate of correctly classified instances from total instances	$\frac{TP+TN}{TP+TN+FP+FN}$
Precision	The classifier's correctness is measured by precision which is the rate of correct predictions	$\frac{TP}{TP+FP}$
Recall	Used to evaluate classifier completeness	$\frac{TP}{TP+FN}$
F-Measure	Weighted average of precision and recall	$2 * \frac{Precision*Recall}{Precision+Recall}$

TABLE IV
EXPERIMENTAL RESULTS OF SIX MACHINE LEARNING ALGORITHMS

Category	Algorithm	Time (s)	Accuracy %	Precision %	Recall %	F-Measure %
Probabilistic	Naïve Bayes (NB)	**0.59**	74.71	68.8	73.6	69.6
	Logistic Regression (LR)	15.87	79.75	73.9	79.8	71.8
Tree	Decision Tree (DT)	113.14	**87.88**	**87**	**86.7**	**85.4**
	Bagging (BG)	65.09	84.12	86.2	84.3	80.1
	AdaBoost (AB)	7.05	79.74	79.7	79.7	81.4
	Random Forest (RF)	**4.74**	**98.80**	**98.4**	**99**	**98.6**

TABLE V
COMPARISON OF OUR EXPERIMENTAL RESULTS WITH PREVIOUS STUDY

Algorithms	Accurcay							
	Naïve Bayes	Logistic Regression	Decision Tree	Bagging	AdaBoost	Random Forest	Gaussian	KNN
[20]	N/A	N/A	71.08	N/A	N/A	82.54	81.38	80.39
[21]	N/A	N/A	N/A	81.04	N/A	88.98	N/A	N/A
Our Results	74.71	79.75	**87.88**	**84.12**	79.74	**98.8**	N/A	N/A

Note: N/A refers to the parameters not considered by the study

forest is the bagging concept, which enhances the overall performance by adding additional randomness to the model while growing the trees, which contributes to achieving high accuracy. Besides the good performance of random forest, it takes the second least time 4.74s to build and run the model. Another high-performance algorithm that belongs to the tree category is the decision tree. It achieved 87.88% accuracy, 87% precision, 86.7% recall, and 85.4% f-measure. The decision tree works on the principle, which iteratively breaks the dataset into two or more sample data. It finds the root node based on the highest entropy value, which helps the tree to choose the most consistent hypothesis among the training set. In contrast to the random forest, the decision tree takes the highest time of 113.14s.

Bagging also significantly performed well. It is a technique that combines the predictions from multiple ML algorithms to predict accurately the severity of the accidents than the individual model. In our experiments, it also works well and shows satisfactory results by achieving 84.12% accuracy followed by 86.2% precision, 84.3% recall, and 80.1% f-measure. Lastly, the AdaBoost algorithm used as an ensemble in ML algorithms also achieved good results using the hyperparameters shown in Table II. It got 79.74% accuracy, 79.7% precision, 79.7% recall, and 81.4% f-measure. In the probabilistic category, both algorithms, the naïve Bayes, and the logistic regression performed efficiently. In this research, the naïve Bayes secured 74.71% accuracy, 68.8% precision, 73.6% recall, and 69.6% f-measure when used with a supervised discretization filter. Finally, the logistic regression achieved 79.75% accuracy, 73.9% precision, 79.8% recall, and 71.8% f-measure.

To further show the effectiveness of our crash prediction framework, we compare the results with the previous research studies. Steven et al. [20] used the same crash dataset published by the Department of Transport UK to predict the crash severity using four machine learning algorithms: Random Forest, Gaussian, Decision Tree, and K-Nearest Neighbor. The performance of these algorithms is evaluated solely relying; on the accuracy metric. Gan et al., [21] predicted the crash severity using machine learning algorithms and their experimental results showed that the Bagging and Random Forest achieved 81.04% and 88.98% accuracy respectively.

The comparison of the results of the study [20], [21] and our study is shown in Table V. It is evident from Table V. that our experiments outperformed their results. Their Random Forest achieved 82.54%, 88.98% and, Decision tree 71.08% accuracies whereas, in our study, we achieved 98.80% and 87.88% accuracies.

V. EXTRACTION OF CRASH PATTERNS

This section aims to analyze the massive traffic dataset and extract the new insightful pattern related to traffic accidents. The authors of this paper believe that the extraction of crash patterns would assist the research community in designing and developing new predictive models. Thus, in turn, would help the road and transport authorities to enhance road safety by preventing and controlling vehicular accidents.

- ***Age Group:*** The analysis shows that the drivers of the age group between 18-30 are the most vulnerable to crashes.

Authorized licensed use limited to: DUBLIN CITY UNIVERSITY. Downloaded on February 03,2024 at 23:13:54 UTC from IEEE Xplore. Restrictions apply.

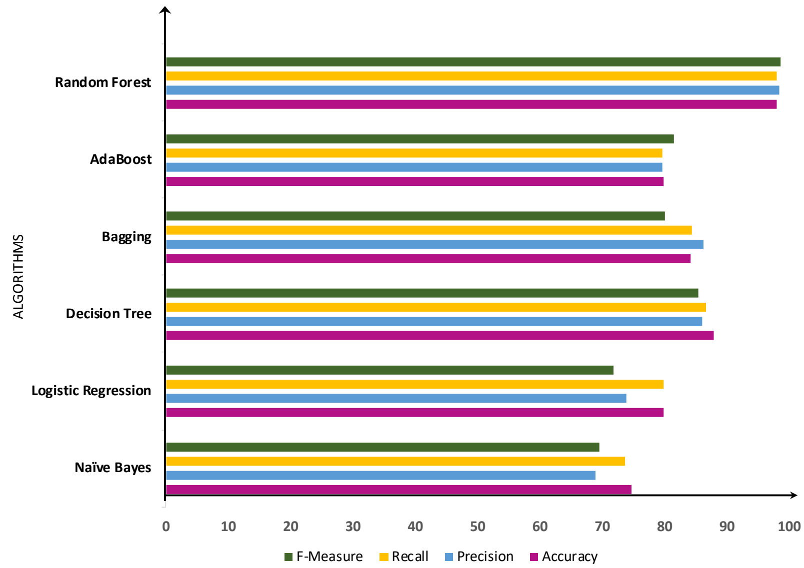

Fig. 2. Performance measures of the developed models

The results; are aligned with real-life as this age group drivers are more irresponsible, usually drive fast, some do not have a license, and lack driving experience. To control road crashes, the government should first educate the people about the traffic rules and regulations and then allow them to come on the roads.

- ***Driver Gender:*** It is a misconception that female drivers commit more accidents as they are not considered expert drivers. Interestingly the analysis shows that male drivers are more involved in accidents than female drivers. Therefore, the results portray that more driving awareness; should be introduced to male drivers to avoid accidents. Besides, it is also noted the most accidents occurred due to the negligence of the driver rather than the passengers or the pedestrians.
- ***Road Type:*** There are different types of roads provided in the dataset such as a roundabout, one-way street, dual carriageway, and single carriageway, etc. The results show that a single carriageway is five times more dangerous than a dual carriageway as 75% of the accidents happened on these roads. Therefore, we suggest that the single carriageway roads should be extended to dual way to lessen accidents, and save precious human lives.
- ***Urban or Rural Area:*** The rural and urban areas have entirely different characteristics concerning the density of road networks, land use, and complex environmental dynamic factors. The analysis reveals that 7% of crashes occurred in the urban area whereas; 6.5% serious accidents happened in the rural area.
- ***Road Surface and Weather Conditions:*** Generally, people think that most accidents happen when the weather is windy or foggy, however, our results divulged that 18.5% of accidents occurred when the weather was perfectly fine, and only 5.75% of accidents happen when the weather is rainy. This is to also note that majority of the accidents happened on a dry surface rather than on wet roads.
- ***Vehicle Manoeuvre:*** A variety of road crashes happen due to vehicle maneuvering. The maneuvers may include reverse, entering or leaving the roadways, taking U-turns, and parking. We analyze that 10.03% of accidents occur during vehicle reversing maneuvers. Similarly, 15%, 1.32%, and 1% of accidents happened due to lane changing, overtaking, and parking maneuvers respectively.
- ***Impact of Time:*** The analysis summarizes that most accidents happened during school timings and some during office hours. Most accidents happen in the morning and evening since the starting and ending times of many institutions are similar. Considering these statistics, either the school and office timings should be altered a bit to prevent accidents, or new roads, intersections and crossings should be constructed on the main rush areas.
- ***Impact of Week Day and Speed Limits:*** The results show that the highest number of accidents happened on Fridays than the weekends. Moreover, more accidents occurred in areas with the speed limit of 20mph, 30mph, and 60mph. The results manifest that the highest number of fatal crashes happened on the roads with 60mph, followed by the roads with 30mph. These insights show that more police officers should be deployed on weekends enforcing people to follow the traffic rules.

VI. Conclusion

In this research, we have developed a prediction framework and analyzed the performance of six state-of-the-art ML algorithms in predicting crash severity in the UK. We have also improved the class imbalance problem of the dataset by using the SMOTE technique. Our experimental results indicate that

Authorized licensed use limited to: DUBLIN CITY UNIVERSITY. Downloaded on February 03,2024 at 23:13:54 UTC from IEEE Xplore. Restrictions apply.

the performance of the crash prediction system is enhanced; with Random Forest, Decision Tree, Bagging, and Logistic regression with 98.80%, 87.88%, 84.12%, and 79.75% accuracies respectively. The findings of this research; demonstrate that normalizing the data and leveraging SMOTE with these algorithms enhanced the effectiveness of the framework in helping road and traffic authorities in predicting the road crash severity accurately. Furthermore, we also analyzed the traffic data and extracted the crucial crash patterns helping the road authorities in implementing and imposing new traffic rules and regulations. In the future, we plan to extend this work by implementing several deep learning, representation learning, and logical reasoning algorithms to provide high predictability and accuracy. Moreover, it should be noted that this research does not focus on optimizing the model hyperparameters, which is also a research direction in the future.

Acknowledgment

This research was funded by the Emirates Center for Mobility Research (ECMR) of the United Arab Emirates University (grant number 31R151).

References

[1] T. K. Bahiru, D. K. Singh, and E. A. Tessfaw, "Comparative study on data mining classification algorithms for predicting road traffic accident severity," in *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*. IEEE, 2018, pp. 1655–1660.

[2] W. H. Organization, *Global status report on alcohol and health 2018*. World Health Organization, 2019.

[3] A. Jamal, M. Zahid, M. Tauhidur Rahman, H. M. Al-Ahmadi, M. Almoshaogeh, D. Farooq, and M. Ahmad, "Injury severity prediction of traffic crashes with ensemble machine learning techniques: a comparative study," *International journal of injury control and safety promotion*, pp. 1–20, 2021.

[4] H. Al Najada and I. Mahgoub, "Big vehicular traffic data mining: Towards accident and congestion prevention," in *2016 International Wireless Communications and Mobile Computing Conference (IWCMC)*. IEEE, 2017, pp. 256–261.

[5] M. Chong, A. Abraham, and M. Paprzycki, "Traffic accident data mining using machine learning paradigms," in *Fourth International Conference on Intelligent Systems Design and Applications (ISDA'04), Hungary*, 2018, pp. 415–420.

[6] H. Al Najada and I. Mahgoub, "Anticipation and alert system of congestion and accidents in vanet using big data analysis for intelligent transportation systems," in *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 2016, pp. 1–8.

[7] A. Elfar, A. Talebpour, and H. S. Mahmassani, "Machine learning approach to short-term traffic congestion prediction in a connected environment," *Transportation Research Record*, vol. 2672, no. 45, pp. 185–195, 2018.

[8] A. Iranitalab and A. Khattak, "Comparison of four statistical and machine learning methods for crash severity prediction," *Accident Analysis & Prevention*, vol. 108, pp. 27–36, 2017.

[9] R. E. AlMamlook, K. M. Kwayu, M. R. Alkasisbeh, and A. A. Frefer, "Comparison of machine learning algorithms for predicting traffic accident severity," in *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)*. IEEE, 2019, pp. 272–276.

[10] H. Ibrahim and B. H. Far, "Data-oriented intelligent transportation systems," in *Proceedings of the 2014 IEEE 15th International Conference on Information Reuse and Integration (IEEE IRI 2014)*. IEEE, 2019, pp. 322–329.

[11] J. Paul, Z. Jahan, K. F. Lateef, M. R. Islam, and S. C. Bakchy, "Prediction of road accident and severity of bangladesh applying machine learning techniques," in *2020 IEEE 8th R10 Humanitarian Technology Conference (R10-HTC)*. IEEE, 2020, pp. 1–6.

[12] "Road safety data - data.gov.uk," https://data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-safety-data, (Accessed on 12/07/2019).

[13] "Attribute-relation file format (arff)," https://www.cs.waikato.ac.nz/ml/weka/arff.html, (Accessed on 09/08/2021).

[14] T. Jayalakshmi and A. Santhakumaran, "Statistical normalization and back propagation for classification," *International Journal of Computer Theory and Engineering*, vol. 3, no. 1, pp. 1793–8201, 2011.

[15] S. A. Arhin and A. Gatiba, "Predicting crash injury severity at unsignalized intersections using support vector machines and naïve bayes classifiers," *Transportation Safety and Environment*, vol. 2, no. 2, pp. 120–132, 2020.

[16] M.-M. Chen and M.-C. Chen, "Modeling road accident severity with comparisons of logistic regression, decision tree and random forest," *Information*, vol. 11, no. 5, p. 270, 2020.

[17] M. S. Satu, S. Ahamed, F. Hossain, T. Akter, and D. M. Farid, "Mining traffic accident data of n5 national highway in bangladesh employing decision trees," in *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*. IEEE, 2017, pp. 722–725.

[18] M. Taamneh, S. Alkheder, and S. Taamneh, "Data-mining techniques for traffic accident modeling and prediction in the united arab emirates," *Journal of Transportation Safety & Security*, vol. 9, no. 2, pp. 146–166, 2017.

[19] N. Kamboozia, M. Ameri, and S. M. Hosseinian, "Statistical analysis and accident prediction models leading to pedestrian injuries and deaths on rural roads in iran," *International journal of injury control and safety promotion*, vol. 27, no. 4, pp. 493–509, 2020.

[20] S. Haynes, P. C. Estin, S. Lazarevski, M. Soosay, and A.-L. Kor, "Data analytics: Factors of traffic accidents in the uk," in *2019 10th International Conference on Dependable Systems, Services and Technologies (DESSERT)*. IEEE, 2019, pp. 120–126.

[21] J. Gan, L. Li, D. Zhang, Z. Yi, and Q. Xiang, "An alternative method for traffic accident severity prediction: using deep forests algorithm," *Journal of advanced transportation*, vol. 2020, 2020.

Authorized licensed use limited to: DUBLIN CITY UNIVERSITY. Downloaded on February 03,2024 at 23:13:54 UTC from IEEE Xplore. Restrictions apply.