Predictive Model for Accident Severity

Solwan M. Mostafa, Sameh A. Salem, and Shahira M. Habashy, *Member, IAENG*

***Abstract*: In the development of sustainable transportation, traffic safety is a significant matter, and predicting the severity of traffic accidents is still a critical problem in the traffic safety field. However, the utilized road traffic accidents (RTAs) datasets suffer from imbalance distribution. This problem leads to a decrease in classification performance, specifically in predicting the minority classes. This paper aims to treat the class imbalance problem through Synthetic Minority Over-sampling Technique (SMOTE), Support Vector Machine-SMOTE (SVM-SMOTE), Borderline-SMOTE(BL-SMOTE), and Adaptive Synthetic (ADASYN), along with proposing an accurate predictive model for accident severity. Different accident severity models are employed, namely Random Forest (RF), K-Nearest Neighbor (KNN), Naïve Bayes (NB) classifiers, Decision Tree (DT), and Extra Trees (ET). These models are tested using real-world datasets. Several evaluation metrics are used to evaluate the proposed model. Experimental results show that the proposed model significantly improves predicting both the minority and majority classes. These results indicate the robustness and reliability of the proposed predictive model in enhancing road traffic safety and management.**

***Index Terms*— Accident severity prediction, ADASYN, borderline-SMOTE, class imbalance problem, decision tree, extra trees, k-nearest neighbor, naïve Bayes, random forest, SMOTE, SVM-SMOTE.**

I. INTRODUCTION

In promoting sustainable transport safety, traffic safety is a significant problem. The impact on society of road accidents will be adverse, including victims, traffic jams and environmental pollution which do not lead to healthy and good growth in the transport system. Several public authorities and the transport sector have developed intelligent transport systems to promote sustainable transport development by increasing the automated information system. Accurate methods are necessary for predicting the severity of traffic accidents to improve traffic safety monitoring and control. In recent years, the rapid progress of science and technology has enhanced new technologies for transport in unprecedented ways. However, there are no precise traffic accident reduction characteristics for these technologies. The World Health Organization (WHO) published Save LIVES- A road safety technical package 2017, stating that more than one million deaths have been recorded in road accidents. Over 50 million people are suffering non-fatal injuries worldwide every year, estimated to be the ninth-largest death cause in the world for each group of age[1]. Road traffic accidents can happen all the time but are predictable and avoidable. Therefore, each traffic research scientist is responsible for reflecting on the causes of accidents in traffic and helping the administration solve the risk of accidents. Over the years, scientists from different backgrounds have explored other models of traffic accident severe study. The modeling analysis is intended to examine the connection between the severity of the accident and its factors, the most commonly used being the discreet Logit or Probit model selection model (e.g.[2],[3],[4],[5],[6]). These results demonstrated that the accurate prediction of traffic accident severity has a valuable role in improving traffic safety management. Since the influential factors of the high-risk segments might be defined to provide helpful ideas for improving road safety, the study's findings will help build or enhance an efficient traffic safety framework within a sustainable transportation system. This framework is crucial for assisting government managers in implementing timely proactive traffic accident prevention strategies and improving road traffic safety.

According to the number of casualties, deaths generally are considerably lower than ownership damage. So, the investigation of the accident severity is significant. Mostly, accident datasets are not balanced that causes a problem in the classification, which is named the class imbalance problem. The class imbalance generally occurs because datasets include occasional occurrences that cause skewness of the used classifier: Thus, as the more the classifiers become accurate in the prediction of the majority class, the less the accuracy in prediction of the minority class [7],[8],[9],[10]. Skewed classifications pose a challenge to predictive modeling since most classifiers assume an equal number of samples for each class. Many research efforts have been dedicated to resolving this problem, yielding many solution approaches. Approaches are broadly classified into three types: 1. techniques of data-level, 2. techniques of algorithm-level, and 3. techniques of cost-sensitive. Techniques of data level are more widely employed because they do not depend on any classifier and can be flexibly combined with other techniques [11]. As a result, these techniques may yield more balanced data than other techniques [12],[13].

The main objective of this paper is to handle the imbalance problem in accident datasets and develop an accurate predictive model for accident severity. The Synthetic Minority Over-Sampling Technique (SMOTE), Support Vector Machine-SMOTE (SVM-SMOTE), Borderline-SMOTE(BL-SMOTE), and Adaptive Synthetic (ADASYN) are adopted for balancing the dataset, which overcomes the minority class classification problem. Comparisons with

Manuscript received June 11, 2021; revised December 25, 2021.
Solwan M. Mostafa is a Teaching Assistant and a Master Student of Computer and Systems Department, Faculty of Engineering, Helwan University, Cairo 11792, Egypt (corresponding author to provide e-mail: solwan_mohamed@h-eng.helwan.edu.eg).
Sameh A. Salem is a Professor of Computer and Systems Department, Faculty of Engineering, Helwan University, Cairo 11792, Egypt (e-mail: sameh_salem@h-eng.helwan.edu.eg).
Shahira M. Habashy is an Associate Professor of Computer and Systems Department, Faculty of Engineering, Helwan University, Cairo 11792, Egypt (e-mail: shahira_heikal@h-eng.helwan.edu.eg).

other competitive prediction methods [14],[15] are carried out. Also, further analysis using 10-folds Cross-Validation (CV) and the Holdout methods are carried out for model training. In this paper, Random Forest (RF), K-Nearest Neighbor (KNN), Naïve Bayes (NB), Decision Tree (DT), and Extra Trees (ET) classifiers are applied. Experimental results show that the proposed predictive model effectively predicts the minority class and enhances the overall classification performance.

The main contribution of this paper is proposing an accurate predictive model for enhancing road traffic safety and management. The proposed model solves the class imbalance problem in datasets and improves classification performances for predicting minority classes. Experimental results show that the proposed model achieves classification accuracies greater than 83%, compared with recently developed models that achieve classification accuracies up to 60% [14]. Also, the results show excellent performance in predicting majority and minority classes compared with [14],[15]. These results indicate the reliability of the proposed model for getting accurate prediction of traffic accident severity which improves traffic safety management.

The paper consists of five sections. Section 1 explores the paper introduction, followed by related work in section 2. Section 3 presents the methodology utilized in this paper, whereas the experimental results are demonstrated in Section 4. Finally, the paper is ended with the conclusion and references.

II. Related Work

As reported by several researchers, several classification algorithms are employed in predicting and analyzing the severity of vehicles' accidents and classifying the patterns of the accidents. Furthermore, we used the great potential of classifiers in preventing and controlling the safety problems related to road accidents.

This section tackles some related works that utilized classification algorithms to predict and analyze road accidents in urban areas, particularly smart cities, to predict and analyze the severity of road traffic accidents (RTAs).

Lécué et al. [16] introduced a Semantic Traffic Analysis while Beshah et al. [17] employed Classification and Adaptive Regression Trees (CART), RF, TreeNet, and hybrid ensemble methodology. According to the comparative results, the Ensemble technique outperformed all other classifiers (single classifiers). The accuracy of TreeNet was 94.54%, CART was 93.52%, and RF was 90.75%, whereas that of ensemble methodology was 95.47%.

Beshah et al. [17] employed different data mining techniques. They used the same dataset size utilized in other research [18],[19]. Still, there were differences in the results, referring to the differences in values of class label that were (injury, non-injury, or fatal) [17], whereas in [18],[19] were (non-injury and injury).

Beshah et al. [18] extended their works using the exact size of dataset 14,254 accident records which included 48 attributes [19]. Their study added TreeNet, CART, and RF to their work [19] for analyzing accident data collected by the Addis Abab traffic office through a tool named " Salford Predictive Miners (SPM)." The results of their work indicated that TreeNet outperformed the other techniques as its accuracy reached 98.94% compared with the other two techniques, which have accuracies (86.59% and 84.5%) for RF and CART, respectively.

We utilized K-Nearest Neighbors, Decision tree (J48), and Naïve Bayes classification techniques for establishing a model capable of predicting and analyzing the factors related to road accident severity. The dataset included 18288 accidents in the city of Addis Ababa. The accuracies of these techniques were 80.8281%, 80.221%, and 79.9967% for K-Nearest Neighbors, Decision tree (J48), and Naïve Bayes, respectively. Another algorithm (PART), with the help of the WEKA tool, is utilized for setting the obtained knowledge in the form of rules [20].

In Hong Kong, Krishnaveni & Hemalatha used a prospective analysis of 34,575 road traffic accidents [21]. They utilized RF, J48, PART, AdaBoostM1, and Naive Bayes classifiers for predicting and detecting the severity of injury and reasons beyond the accidents. To obtain minimal dimensionality of the Accident dataset, they used Genetic Algorithm (GA) to select the features. RF outperforms all other algorithms according to the comparative results of the classifiers. No percentages are obtained.

Regarding the real-time analysis of the data related to the accidents, a Real-time Transportation Data Mining (RTransDmin) technique is proposed by [22]. This technique is also able to predict information related to traffic accidents. In the same study, Decision tree ADTree and J48 algorithms with the help of WEKA and DTREG tools are utilized for building a model related to a dataset that included 1385 accident records. This dataset is obtained from the department of transport in England. Confusion matrices resulting from the DTERG are characterized by accuracies 87.2 and 85.9 % for training and testing datasets, respectively, and the scatter plot is obtained by WEKA.

The study of Perone [23] is one of the significant studies used to evaluate the injury severity through utilizing Naive Bayes, RF, Support Vector Machine (SVM), LR, and KNN, for constructing the prediction models. The used dataset included 20798 accident records obtained from the traffic department in Porto Alegre/RS (Brazil). Pandas library is used to analyze data, whereas Scikit-learn framework is utilized for preprocessing operations. The results of Area Under the ROC Curve (AUC) indicated that the best-obtained scores are 0.94 for both LR and SVM. However, the RF, KNN, and Naïve Bayes scores are 0.93, 0.9, and 0.83, respectively. F-measure is also employed for evaluating the used algorithms. The results obtained indicated the same measures for LR and SVM (0.89), (0.88) for RF, (0.85) for KNN, and (0.43) for Naïve Bayes. Some criticisms are directed to this study, as the used dataset does not include information about drivers, victims, and vehicles, and thus the author did not use feature selection techniques.

The SVM, Gaussian RBF kernel, and the polynomial kernel investigated driver injury severity patterns. A two-year dataset of accidents in New Mexico is collected and used in this study. The study results indicated that reasonable prediction is obtained from applying the SVM model, whereas the performance of the polynomial kernel algorithm is better than the Gaussian RBF Kernel [24].

RF, Artificial Neural Network (ANN), and SVM are employed for predicting the severity of accidents in the UK. The used dataset size was 79751 records in 2014. The ANN with an accuracy of 61.4 %, and then SVM with an accuracy of 54.8% [25]. The superiority of RF is confirmed by S.Ramya et al. [26].

Haynes et al. [15] used an imbalanced dataset from the UK to find the factors affecting the number of accidents and the accompanied mortalities. They used RF, KNN, Decision Tree, and Gaussian for predicting the accidents' severity. The results indicated that RF outperformed the other techniques. This study is used as a base for comparing the results of our current study.

Fiorentini et al. [14] focused on exploring the influence of a balancing technique, namely random under-sampling the majority class (RUMS), to solve the class imbalance problem in datasets and improve the performance of the classifiers in predicting the minor classes. Classifiers models such as the random tree, KNN, LR, and RF are used in this study. The study results indicated that attained accuracies were very low (~ <60%). Also, this study is used as a base for comparing the results of our current study.

Crash accidents are implemented on a dataset in Ghana through the employment of PART, MLP, and SimpleCART to evaluate the classifiers' performance and determine the significant factors for the crash of motorcycles. Weka tools are used for comparing and analyzing datasets. For selecting the most influential factors, the InfoGainAttributeEval technique is employed. The highest performance is obtained through the SimpleCART classifier [27].

III. THE METHODOLOGY

This section presents the block diagram of the proposed predictive model phases, as shown in Fig. 1.

A. Data Splitting Phase

This paper employed 10-folds Cross-Validation and Holdout (training data 70% and testing data 30%) to break these datasets into training and testing datasets.

In the 10-folds CV method, the dataset is separated into ten distinct subsets. We used nine subsets to build the predictive model and quit the tenth subset as test data. The model is averaged against each of the folds. In the holdout method, the dataset is separated into only two subsets.

B. Balancing Phase

Many techniques have been suggested in recent years to address the weak performance of imbalanced datasets [28]. They are divided into three main categories: 1. techniques of data-level (resampling). 2. techniques of algorithm-level. 3. techniques of cost-sensitive.

1. *Data-Level Techniques (Resampling)*

These techniques adjust the training samples to balance the class distribution, enabling classifiers to work similarly to standard classification. The three principal techniques used to handle imbalanced datasets are under-sampling, over-sampling, and synthetic data generation (e.g., synthetic minority oversampling). These techniques have the advantage of being independent of any classification algorithm. On the other hand, one of the most significant disadvantages of these techniques is deciding how much sampling to employ. The over-sampling technique involves adding samples to the minority class. These samples can be produced by copying the minority class samples or generating new synthetic samples.

In contrast, the exact copies of the sample cause the over-fitting problem. The under-sampling technique involves randomly removing the majority class samples. These lead to the loss of valuable information, so we must select the under-sampling removing percentage carefully [29]. More sophisticated techniques have been developed to solve these problems, e.g. (SMOTE, BL-SMOTE, SVM-SMOTE, and ADASYN)

SMOTE [12] is an oversampling technique in which synthetic samples are created for the minority class. It creates new samples of the minority class by randomly interpolating pairs of the closest neighbors in the minority class. The steps of this technique are shown in Fig. 2.

Borderline-SMOTE [13] is a modification in which samples of the minority class far from the majority class boundary may contribute less to the classifier than samples near the border. As a result, the Borderline-SMOTE preferentially generates synthetic samples along the decision border. Borderline is a region where the samples of minority classes are near the majority. The steps of this technique are shown in Fig. 3.

SMOTE-SVM [29] generates new samples near the decision boundary. SMOTE-SVM, like borderline-SMOTE, considers that the decision border is the optimum area to generate new samples. SMOTE-SVM employed support vectors to detect decision boundaries. The steps of this technique are shown in Fig. 4.

ADASYN [30] sets weight for each minority sample based on their learning difficulties. The distribution of each sample is used to estimate the number of synthetic samples required. We generate extra data points for the hard-to-learn minority samples. ADASYN adjusts the decision boundary to let the classifier focus on the hard-to-learn data. The steps of this technique are shown in Fig. 5.

In the balancing phase, SMOTE, borderline-1-SMOTE, SVM-SMOTE, and ADASYN are adopted due to their advantages.

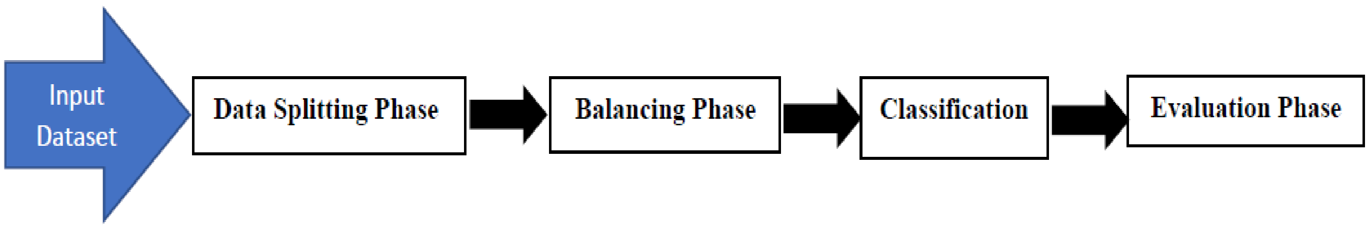


Fig. 1. The Block Diagram of the Proposed Predictive Model

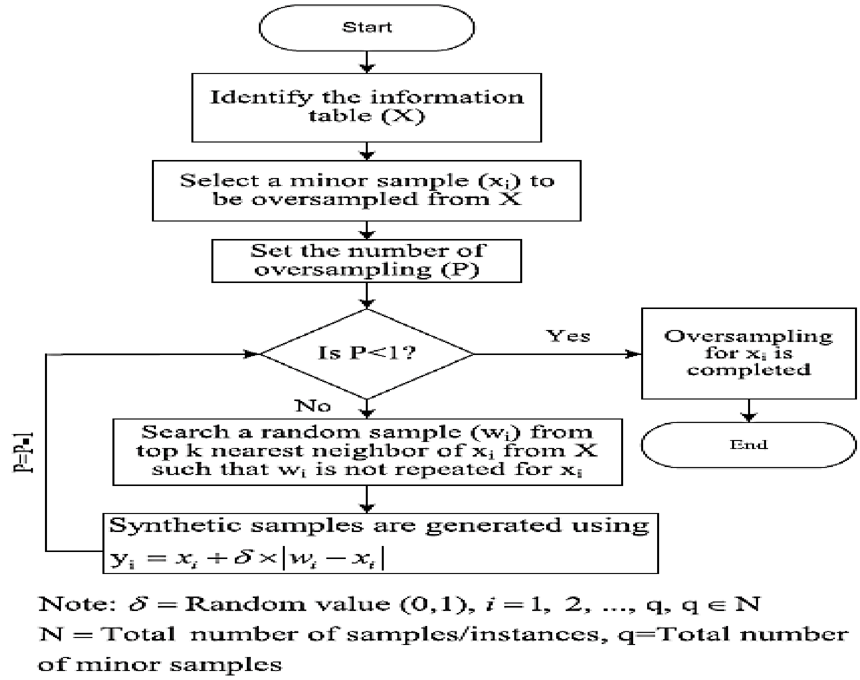


Fig. 2. The Flow Chart of SMOTE Technique

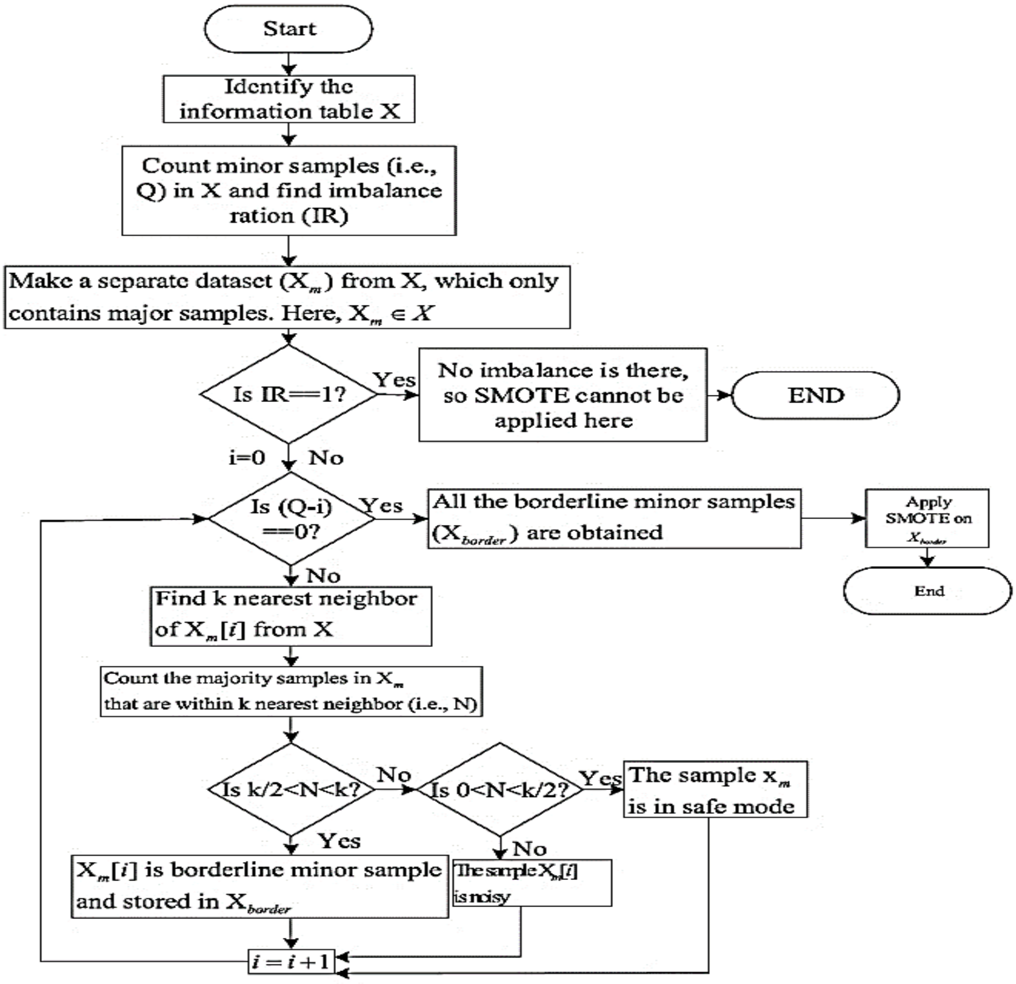
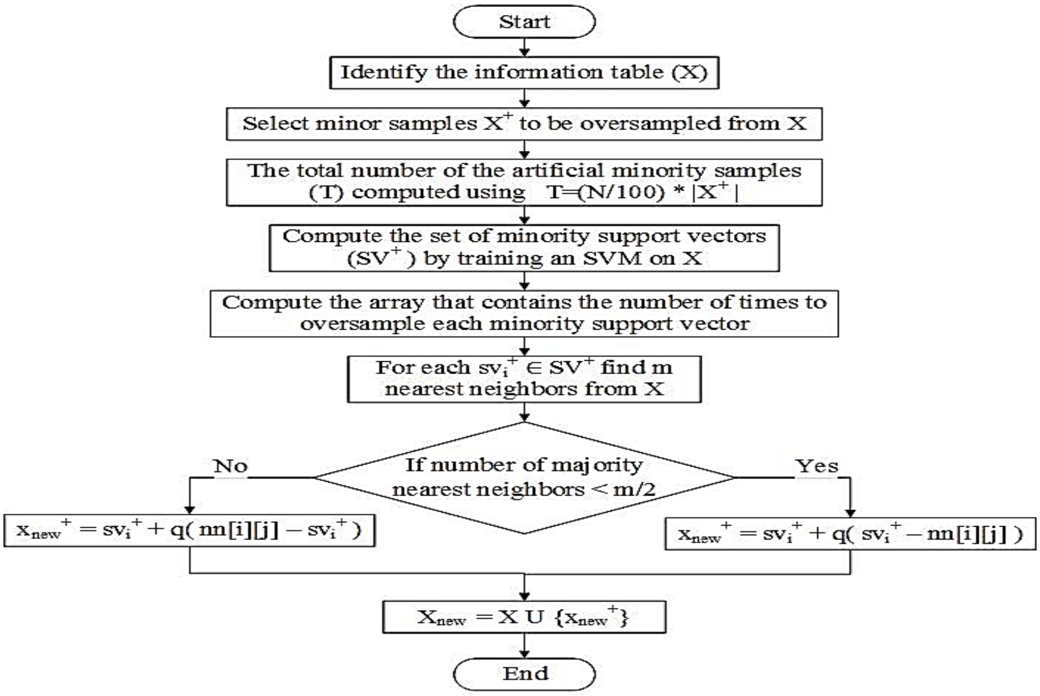
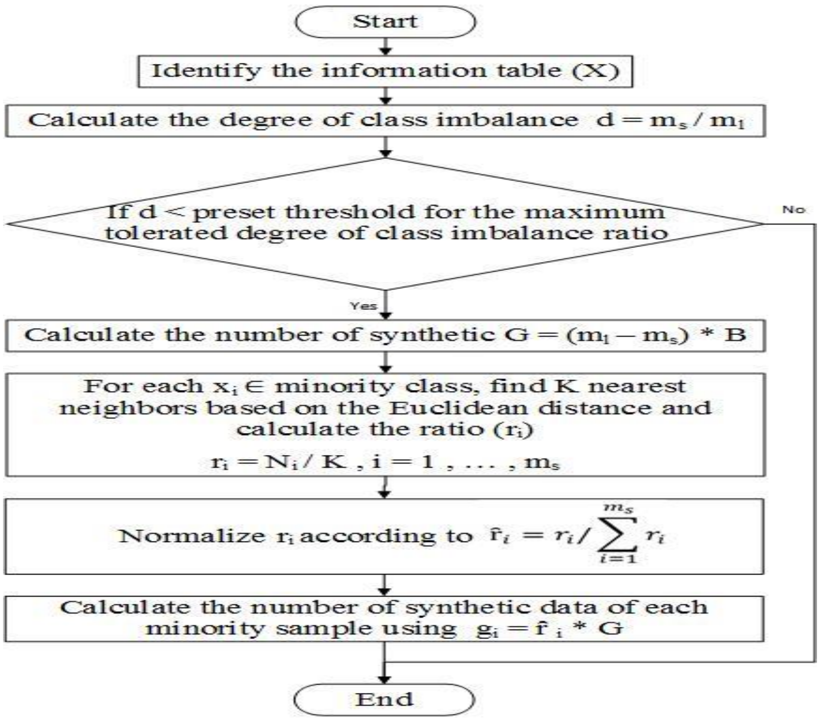


Fig. 3. The Flow Chart of BL-SMOTE Technique



Note: N: Over-sampling level.
m: Number of nearest neighbors used to decide how to over-sample.
Xnew: Over-sampled training set.
nn: Array that contains k minority nearest neighbors of each minority support vector.
q: random number in [0, 1].

Fig. 4. The Flow Chart of SVM-SMOTE Technique



Note: m_s :the number of minority class samples.
m_l :the number of majority class sample.
N_i :the number of samples in the K nearest neighbors of each sample that belong to the majority class.
B : the desired balance level.

Fig. 5. The Flow Chart of ADASYN Technique

2. Algorithm-Level Techniques

They focused on adapting the classification algorithms approach to be more attuned to class imbalance problems. They are non-flexible and reliant on the classifier [31].

3. Cost-Sensitive Techniques.

These techniques assess the cost of misclassifying samples. They do not result in a balanced distribution of data. These techniques use cost matrices to highlight the imbalanced learning problem, which illustrates misclassification cost. The cost of misclassification is unknown from the data; therefore, defining costs is problematic [32],[33].

C. Classification Phase

Supervised classification algorithms aim to use just training data to separate problem classes (with as wide a margin as possible) [34].

1. *Random Forest Classifier*

Breiman [35] is considered the first to propose the random forest classifier. This classifier is utilized as a data mining tool for solving problems related to classification and regression. The accuracy of the classifier is increased when the classification is determined using voting. Also, growing an ensemble of trees leads to an increase in the accuracy of the classifier. Random vectors are created to grow these ensembles. Each random vector generates one tree. For classification and regression, RF is composed of trees. By canvassing the outcomes of trees, the class of the sample is determined. Since over-fitting does not occur in large RFs, the generalization error merges to a limiting value when adding more trees to the RF. The higher accuracy requires the availability of low skewness and correlation. It is necessary to have no clipping for trees and randomize the variables in each node to obtain low skewness and correlation [36],[37].

2. K- Nearest Neighbor Classifier

Cover and Hart are the first to propose the KNN classifier. It is a machine learning algorithm called a lazy learning algorithm because of its low computational cost and ease of implementation. It attempts to classify a sample by examining the nearest k samples in the feature space. The majority class of the k closest samples belongs to is chosen as the class of the new sample. As a result, KNN assigns the class of the k nearest set of previously classified points to an unclassified sample point [14],[38].

3. *Naive Bayes Classifier*

The Naive Bayes (NB) classifier is a Bayesian network specifically developed for classification problems [39]. This simple probabilistic classification model computes the likelihood of a class variable given specific instances of feature variables and then predicts the class of the class variable with the highest posterior probability. This computation is effectively conducted by making the strong independence assumption that all of the feature variables are conditionally independent given the value of the class variable. The fundamental advantage of NB over other machine learning models is its computational simplicity.

4. The Decision Tree Classifier

The Decision Tree (DT) Classifier divides observed data into mutually exclusive categories and then uses that tree structure to make decisions. Each node is split into child nodes to build the tree, starting with the root node. It is possible to split DT algorithms using different criteria. The most widely used decision tree algorithms are ID3, CART, and C4.5, which use impurity-based splitting criteria to reduce impurities [40].

5. The Extra Trees Classifier

The Extra Trees (ET) classifier refers to the extremely randomized trees classifier [41]. The extra trees classifier relies on building multiple unpruned top-down decision trees and integrating the results. A majority vote solves the classification problem. ET is a Random Forest modified version. Unlike RF, which uses bootstrap sampling, Extra Trees uses the entire training sample to build the trees. The ET classifier also differs from Random Forest in that the cut-point for each tree node is estimated randomly, independent of the target feature. K number features are randomly selected to build randomized trees. The optimum value of K depends upon the problem characteristics and the percentage of irrelevant features that decrease the model's performance. ET may reduce model variance slightly more than RF in terms of overfitting.

D. Evaluation Phase

The reason for using multiple metrics is that accuracy does not always express the classifier's optimal performance, especially in imbalanced data. This problem is called the accuracy paradox problem. Accuracy is not considered a good metric because data imbalance may cause classifier accuracy to appear high, although there is a mistake in classifying the minor classes. Precision and recall are better measures in such cases [42],[43].

The following metrics are used for evaluating the model performance: False Positive Rate (FPR), accuracy, True Positive Rate (TPR) or recall, Area Under the ROC Curve (AUC), F-measure, Kappa statistic, Matthews Correlation Coefficient (MCC), and the training time of the model.

The objective of the classifier's accuracy is to represent the overall performance of the classifier [44]. The ratio of False Positive with respect to all negative samples is expressed by FPR [45]. The TPR or recall can be defined as the ratio of positive samples detected correctly [44]. The precision indicates the goodness of positive predictions [44]. The F-measure is expressed by the harmonic mean of both precision and recall, and its significant advantage is its ability to combine two metrics into a more compact one [44]. AUC is obtained by calculating the area under the ROC (Receiver Operating Characteristic) curve. This curve results from the relation between FPR and TPR [44]. The essential advantage of AUC is to judge the capability of the classifier to discriminate between classes. The Kappa statistic measures the degree of inter-rater agreement. Its value is estimated according to Table I. The Interpretation of agreement degree [45] is illustrated in Table II. MCC measures the differences between actual values and predicted values [46]. It can be defined in terms of a confusion matrix C for K classes. Finally, we calculated the time spent in building the model of each classifier.

TABLE I
2×2 CONTINGENCY TABLE

RATER 1	RATER 2		TOTAL
	1	2	
1	P_{11}	P_{12}	P_1
2	P_{21}	P_{22}	P_2
TOTAL	P_1	P_2	1

TABLE II
KAPPA STATISTIC INTERPRETATION [45]

KAPPA Statistic	Degree of Agreement
< 0.00	Poor-Degree
0.00-0.20	Slight-Degree
0.21-0.40	Fair-Degree
0.41-0.60	Moderate-Degree
0.61-0.80	Substantial-Degree
0.81-1.00	Almost perfect-Degree

Equations (1) – (7) express the accuracy, FPR, TPR or the recall, the precision, F-measure, Kappa statistic, and MCC.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \quad (1)$$

$$\text{FPR} = \frac{\text{FP}}{\text{TN} + \text{FP}} \quad (2)$$

$$\text{TPR(Recall)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

$$\text{Precision} = \frac{TP}{\text{TP} + \text{FP}} \quad (4)$$

$$F\_measure = \frac{2 * (\text{Recall} * \text{Precision})}{\text{Recall} + \text{Precision}} \quad (5)$$

$$\text{Kappa statistic} = \frac{(\text{p}_0 - \text{p}_e)}{(1 - \text{p}_e)} \quad (6)$$

$$MCC = \frac{\text{c} * \text{s} - \sum_k^K p_k * t_k}{\sqrt{(s^2 - \sum_k^K p_k^2)(s^2 - \sum_k^K t_k^2)}} \quad (7)$$

where:

- TP (**True Positive**): expresses the number of correctly classified positive samples;
- TN (**True Negative**): expresses the number of correctly classified negative samples;
- FP (**False Positive**): denotes the number of false predictions negative samples;
- FN (**False Negative**): denotes the number of false predictions positive samples.
- $p_0 = p_{11} + p_{22}$ = (TP + TN) / (TP + FN + FP + TN)
- $p_e = p_1 p_1 + p_2\, p_2$ with $p_1 p_1$ = (TP + FN) (TP + FP) /(TP + FN + FP + TN)^2 $p_2\, p_2$ = (FP + TN) (FN + TN) / (TP + FN + FP + TN)^2
- c= $\sum_i^K \sum_j^K C_{ij}$: the total number of elements correctly predicted
- s= $\sum_k^K C_{kk}$: the total number of elements
- $p_k = \sum_i^K C_{ki}$: the number of times that class k was predicted
- $t_k = \sum_i^K C_{ik}$: the number of times that class k truly occurred (row total)

IV. RESULTS AND DISCUSSIONS

This section presents the utilized real-world datasets and discusses the experiments and the results of the proposed predictive model.

The experiments are performed using two real-world datasets. The obtained results are compared with the obtained results in [14],[15]. [15] employed imbalanced training data to train the classifiers, then added some features to enhance the classifier's performance. [14] employed imbalanced training data to train the classifiers and then used RUMS-based training data. Different comparisons and analyses are discussed to show the performance of the proposed predictive model in predicting traffic accident severity.

A. *Real-World Datasets*

Two real-world datasets are used in this paper [47]. The first one is the dataset for RTAs in the UK during 2016, with a total number of accidents of 136,621. The second one is the dataset for RTAs in York, Great Britain, from 2005 to 2018, with a total number of accidents of 6515.

We chose only the most effective RTA features that can be used to train predictive models to learn patterns and thus predict the severity of accidents based on RTA historical data, as demonstrated experimentally in [14],[15]. Table III shows the features used in building the predictive model in [15]. Table IV shows other additional features used to build the predictive model in [14]. Table V shows the class label description, the number of instances belonging to each class, and its percentage in the dataset.

TABLE III
DATASET-1 FEATURES

Feature	Value	Description
Road surface conditions	1	Dry
	2	Wet or damp
	3	Snow
	4	Frost or ice
	5	Flood over 3cm. deep
	6	Oil or diesel
	7	Mud
	-1	Data missing or out of range
Weather conditions	1	Fine no high winds
	2	Raining no high winds
	3	Snowing no high winds
	4	Fine + high winds
	5	Raining + high winds
	6	Snowing + high winds
	7	Fog or mist
	8	Other
	9	Unknown
	-1	Data missing or out of range
Light conditions	1	Daylight
	4	Darkness - lights lit
	5	Darkness - lights unlit
	6	Darkness - no lighting
	7	Darkness - lighting unknown
	-1	Data missing or out of range

TABLE III
DATASET-1 FEATURES CONTINUED

Feature	Value	Description
Day of week	1	Sunday
	2	Monday
	3	Tuesday
	4	Wednesday
	5	Thursday
	6	Friday
	7	Saturday
1st road class	1	Motorway
	2	A(M)
	3	A
	4	B
	5	C
	6	Unclassified
Urban or rural	1	Urban
	2	Rural
	3	Unallocated
Speed limit	20	20 MPH
	30	30 MPH
	40	40 MPH
	50	50 MPH
	60	60 MPH
	70	70 MPH
	-1	Data missing or out of range
Number of vehicles	Ex: 1,2,3…	Vehicles number in the accident
Time	Ex: 8:15	The time when the accident occurred
Longitude	Ex: -0.279	The coordinate where the accident occurred
Latitude	Ex: 51.584	The coordinate where the accident occurred

TABLE IV
ADDITIONAL FEATURES OF DATASET-2

Feature	Value	Description
Junction Detail	0	Not at junction or within 20 meters
	1	Roundabout
	2	Mini-roundabout
	3	T or staggered junction
	5	Slip road
	6	Crossroads
	7	More than 4 arms (not roundabout)
	8	Private drive or entrance
	9	Another junction
	-1	Data missing or out of range
Pedestrian Crossing-Human Control	0	None within 50 meters
	1	Control by school crossing patrol
	2	Control by another authorized person
	-1	Data missing or out of range
Pedestrian Crossing-Physical Facilities	0	No physical crossing facilities within 50 m
	1	Zebra
	4	Pelican, puffin, toucan, or similar non junction pedestrian light crossing
	5	Pedestrian phase at traffic signal junction
	7	Footbridge or subway
	8	Central refuge
	-1	Data missing or out of range
Special conditions at the site	0	None
	1	Auto traffic signal-out
	2	Auto signal part defective
	3	Road sign or marking defective or obscured
	4	Roadworks
	5	Road surface defective
	6	Oil or diesel
	7	Mud
	-1	Data missing or out of range
Carriageway Hazards	0	None
	1	Vehicle load on road
	2	Another object on road
	3	Previous accident
	4	Dog on road
	5	Another animal on road
	6	Pedestrian in carriageway - not injured
	7	Any animal in carriageway (except ridden horse)
	-1	Data missing or out of range
2nd Road Class	0	Not at junction or within 20 meters
	1	Motorway
	2	A(M)
	3	A
	4	B
	5	C
	6	Unclassified
Road Type	1	Roundabout
	2	One way street
	3	Dual carriageway
	6	Single carriageway
	7	Slip road
	9	Unknown
	12	One way street/ slip road
	-1	Data missing or out of range
Number of casualties	Ex: 1,2,3…	Casualties number in the accident

TABLE V
CLASS LABEL DESCRIPTION

Accident Severity	Value	Number of instances in dataset-1	Number of instances in dataset-2
Fatal (Class1)	1	1695 (1.24%)	64(0.98%)
Serious (Class2)	2	21,725 (15.9%)	855(13.12%)
Slight (Class3)	3	113,201 (82.85%)	5593(85.85%)
		Total= 136,621(100%)	Total= 6515(100%)

B. *Experiments and Results*

After splitting the datasets, they are analyzed using different classifiers, RF, KNN, NB, DT, and ET, to predict accident severity. In this experiment, the classifiers are used on the two datasets separately. The datasets are balanced using oversampling techniques, SMOTE, SVMSMOTE, BLSMOTE, and ADASYN. Table VI reports the evaluation results (accuracy, training time, and Kappa statistics) of the proposed predictive model using different classifiers and data splitting methods. In all the cases, the proposed predictive model-SMOTE-RF (10-folds CV) performs better than the others based on accuracy and Kappa statistics. The KNN and NB are the fastest during training the model. ET has moderate accuracy, but it is very slow to learn. DT and NB perform worse than the others based on accuracy and Kappa statistics. Dataset-1 is bigger than dataset 2. This criterion leads to an increase in models' performance. Table VII shows the evaluation results of each class and a comparison for the proposed predictive models with RF classifier using Dataset 1. It shows that the imbalanced data train the RF classifier after adding additional features as recorded in [15] the balanced data train the RF. [15] employed holdout (70%, 30%) for data splitting, whereas we employed holdout (70%, 30%) and 10-folds CV. From the confusion matrix in [15], TRP, FPR, precision, and F-measure are computed to compare the result. Although the accuracy of the Imbalanced dataset-RF is moderate, the results of TPR, precision, and F-Measure of Imbalanced dataset-RF are very low for minor classes (fatal and serious) and bias to the major class (slight). This bias is because the distribution of the original dataset is imbalanced. The accuracy paradox problem occurs when the classifier is trained using an imbalanced dataset. After oversampling the dataset using different techniques, SMOTE, SVM-SMOTE, borderline1-SMOTE, and ADASYN, the RF classifier achieves a significant enhancement in predicting all classes and increasing the accuracy as shown in Table VII. In comparing those models, the proposed model-SMOTE-RF obtained superiority according to the evaluation metrics used.

TABLE VI

EVALUATION RESULTS OF THE PROPOSED PREDICTIVE MODELS USING DIFFERENT CLASSIFIERS

Oversampling	Data Splitting	Classifier	Accuracy		Training Time		Kappa Statistics	
			Dataset-1	Dataset-2	Dataset-1	Dataset-2	Dataset-1	Dataset-2
SMOTE	10-folds CV	RF	**92.80 %**	**86.04%**	69.44 Sec	3.65 Sec	**0.890**	**0.7904**
		KNN	90.14 %	81.60 %	0.09 Sec	**0.01 Sec**	0.8512	0.7249
		NB	77.25 %	67.82 %	0.04 Sec	0.06 Sec	0.6464	0.5170
		DT	73.55%	77.80%	5.14 Sec	0.70 Sec	0.6032	0.6672
		ET	84.66%	81.81%	78.71 Sec	24 Sec	0.7700	0.7273
	Holdout	RF	92.60%	84.90%	78.22 Sec	4.57 Sec	0.8879	0.7735
		KNN	90.18 %	80.42%	0.07 Sec	0.04 Sec	0.8518	0.7076
		NB	76.70 %	67.47 %	0.04 Sec	**0.01 Sec**	0.6384	0.5115
		DT	73.69%	80.00%	0.51 Sec	0.08 Sec	0.6053	0.6702
		ET	84.62%	82.32%	8.78 Sec	2.21 Sec	0.7693	0.7350
SVMSMOTE	10-folds CV	RF	85 %	82.85 %	102.60 Sec	5.79 Sec	0.7646	0.7358
		KNN	78.44 %	80.24 %	**0.02 Sec**	**0.01 Sec**	0.6703	0.697
		NB	66.10 %	67.89 %	0.44 Sec	**0.01 Sec**	0.4879	0.5059
		DT	77.34%	79.79%	4.96 Sec	0.56 Sec	0.6501	0.6886
		ET	85.86%	84.08%	65.13 Sec	19.87 Sec	0.7812	0.7542
	Holdout	RF	82.77%	82.11%	44.46 Sec	4.12 Sec	0.7415	0.725
		KNN	77.51 %	79.64 %	0.06 Sec	0.02 Sec	0.6563	0.688
		NB	66.34 %	67.33 %	0.03 Sec	**0.01 Sec**	0.4896	0.4998
		DT	77.25%	80.00%	0.47 Sec	0.06 Sec	0.6486	0.6919
		ET	85.72%	84.12%	5.80 Sec	1.71 Sec	0.7791	0.7550
BLSMOTE	10-folds CV	RF	86.08%	82.08 %	114.79 Sec	5.58 Sec	0.7912	0.7311
		KNN	79.02%	79.47 %	0.04 Sec	**0.01 Sec**	0.6853	0.692
		NB	66.54 %	62.05%	0.05 Sec	**0.01 Sec**	0.4927	0.4307
		DT	78.28%	79.50%	5.24Sec	1.19 Sec	0.6743	0.6928
		ET	86.63%	82.94	64.53Sec	21.22 Sec	0.7995	0.7442
	Holdout	RF	85.44%	80.64 %	52.51 Sec	5.1 Sec	0.7817	0.7096
		KNN	77.29 %	78.49 %	0.04 Sec	**0.01 Sec**	0.6594	0.6776
		NB	64.26 %	62.05%	0.05 Sec	**0.01 Sec**	0.4646	0.4307
		DT	78.16%	79.54%	0.52 Sec	0.07 Sec	0.6725	0.6934
		ET	86.87%	83.20%	7.31 Sec	2.53 Sec	0.8031	0.7481
ADASYN	10-folds CV	RF	85.29%	82.85%	91.68 Sec	18.29 Sec	0.779	0.7429
		KNN	75.15%	77.68%	9.17 Sec	5.87 Sec	0.626	0.6657
		NB	51.63%	40.82%	0.44 Sec	0.11 Sec	0.274	0.1064
		DT	76.18%	79.55%	5.40 Sec	1.01 Sec	0.6428	0.6935
		ET	86.35%	83.32%	77.39 Sec	24.11 Sec	0.7951	0.7501
	Holdout	RF	84.67%	82.61%	5.27 Sec	0.92 Sec	0.7699	0.7393
		KNN	75.14%	74.73%	0.45 Sec	**0.01 Sec**	0.6260	0.6220
		NB	51.63%	40.83%	0.03 Sec	**0.01 Sec**	0.2743	0.1064
		DT	76.11%	80.20%	0.62 Sec	0.06 Sec	0.6414	0.7034
		ET	86.48%	83.63%	7.31 Sec	1.97 Sec	0.7970	0.7545

TABLE VII

EVALUATION RESULTS AND COMPARISON FOR THE PROPOSED PREDICTIVE MODELS WITH THE RF CLASSIFIER -DATASET 1

Data Splitting			Holdout				10-folds CV			
			Proposed model-oversampling technique X-RF							
Metric	Class	Imbalanced dataset [15]	SMOTE	SVMSMOTE	BLSMOTE	ADASYN	SMOTE	SVMSMOTE	BLSMOTE	ADASYN
TRP (Recall)	Fatal	0.152	**0.993**	0.974	0.989	0.984	**0.993**	0.970	0.988	0.985
	Serious	0.039	**0.862**	0.764	0.798	0.817	0.859	0.761	0.810	0.822
	Slight	**0.957**	0.914	0.866	0.777	0.741	0.922	0.871	0.786	0.756
FPR	Fatal	0.012	**0.004**	0.016	0.017	0.031	**0.004**	0.014	0.015	0.029
	Serious	0.124	0.043	0.078	0.107	0.121	**0.039**	0.078	0.104	0.115
	Slight	0.467	**0.065**	0.151	0.095	0.081	0.067	0.150	0.091	0.080
Precision	Fatal	0.151	0.991	0.944	0.967	0.938	**0.992**	0.951	0.970	0.943
	Serious	0.220	0.885	0.864	0.787	0.783	**0.895**	0.860	0.795	0.793
	Slight	**0.962**	0.898	0.789	0.805	0.818	0.896	0.782	0.815	0.824
F-Measure	Fatal	0.151	**0.992**	0.959	0.978	0.960	**0.992**	0.961	0.979	0.964
	Serious	0.066	**0.873**	0.811	0.793	0.800	**0.873**	0.810	0.802	0.807
	Slight	**0.959**	0.906	0.825	0.791	0.777	0.906	0.822	0.800	0.789
AUC	Average	NA.	0.976	**0.994**	0.954	0.961	0.976	0.993	0.957	0.954
MCC	Average	NA.	0.888	0.769	0.781	0.769	**0.891**	0.770	0.791	0.800
Accuracy		85.08%	92.60 %	84.93%	85.44%	84.56%	**92.80 %**	85.10%	86.08%	85.29%

The single metric used in [15] is the accuracy when using the KNN classifier. The data splitting approach is holdout (70%,30%). Table VIII shows only the accuracy of the KNN classifier that is trained by the imbalanced training data as recorded in [15] then trained by the balanced data. Also, the proposed model-SMOTE-KNN(Holdout), seems to increase the model's accuracy from 80.39% to 90.18%.

By observing the specific performance for each class in Table IX, we can confirm that the proposed model-SMOTE-KNN is strong in predicting all classes, whether the data is split using 10-folds CV or holdout during the training phase. Although FPR values of the imbalanced dataset for fatal and serious accidents are higher than others, the result is very close to the proposed model-SMOTE-KNN. There is an excellent enhancement in FPR for slight accidents. Tables X, XI, and XII show the evaluation results of each class for the proposed predictive model with the NB, DT, and ET classifiers using dataset 1, respectively. The results indicate that all models enhance predicting each class, and there is no superiority based on all used metrics between them.

TABLE VIII

THE ACCURACY RESULTS AND COMPARISON FOR THE PROPOSED PREDICTIVE MODELS WITH THE KNN CLASSIFIER -DATASET 1

Metric	Accuracy
Imbalanced dataset-KNN[15] (Holdout)	80.39%
Proposed model-SMOTE-KNN(Holdout)	**90.18%**
Proposed model-SMOTE-KNN (10-folds CV)	90.14 %
Proposed model-SVMSMOTE-KNN (Holdout)	77.51 %
Proposed model-SVMSMOTE-KNN (10-folds CV)	78.44 %
Proposed model-BLSMOTE-KNN (Holdout)	77.29 %
Proposed model-BLSMOTE-KNN (10-folds CV)	79.02%
Proposed model-ADASYN-KNN (Holdout)	75.14%
Proposed model-ADASYN-KNN (10-folds CV)	75.15%

TABLE IX

EVALUATION RESULTS FOR THE PROPOSED PREDICTIVE MODELS WITH THE KNN CLASSIFIER -DATASET 1

Data Splitting			Holdout				10-folds CV			
			Proposed model-oversampling technique X-KNN							
Metric	Class	Imbalanced dataset	SMOTE	SVMSMOTE	BLSMOTE	ADASYN	SMOTE	SVMSMOTE	BLSMOTE	ADASYN
TRP (Recall)	Fatal	0.017	**0.990**	0.973	0.982	0.989	0.989	0.967	0.986	**0.990**
	Serious	0.063	0.860	0.763	0.742	0.871	0.858	0.771	0.782	**0.911**
	Slight	**0.932**	0.855	0.677	0.597	0.391	0.856	0.695	0.602	0.498
FPR	Fatal	**0.004**	0.008	0.065	0.089	0.113	0.006	0.062	0.071	0.081
	Serious	**0.065**	0.073	0.166	0.167	0.237	0.074	0.156	0.165	0.198
	Slight	0.931	0.065	0.120	0.085	0.026	0.067	0.118	0.079	**0.025**
Precision	Fatal	0.071	0.984	0.806	0.844	0.806	**0.988**	0.813	0.875	0.855
	Serious	0.170	**0.820**	0.749	0.695	0.662	0.819	0.760	0.703	0.711
	Slight	0.809	0.891	0.783	0.777	0.883	0.888	0.790	0.792	**0.910**
F-Measure	Fatal	0.028	0.987	0.881	0.908	0.888	**0.988**	0.884	0.927	0.918
	Serious	0.092	**0.839**	0.756	0.718	0.752	0.838	0.766	0.741	0.798
	Slight	0.866	**0.873**	0.726	0.675	0.543	0.872	0.740	0.684	0.644
AUC	Average	0.555	0.927	0.915	0.908	**0.961**	0.926	0.920	0.919	0.924
MCC	Average	0.008	**0.851**	0.674	0.660	0.651	0.850	0.687	0.687	0.719

TABLE X
EVALUATION RESULTS FOR THE PROPOSED PREDICTIVE MODELS WITH THE NB CLASSIFIER -DATASET 1

Data Splitting			Holdout				10-folds CV			
			Proposed model-oversampling technique X-NB							
Metric	Class	Imbalanced dataset	SMOTE	SVMSMOTE	BLSMOTE	ADASYN	SMOTE	SVMSMOTE	BLSMOTE	ADASYN
TRP (Recall)	Fatal	0.164	0.834	0.802	**0.862**	0.697	0.854	0.824	0.808	0.687
	Serious	0.103	0.415	**0.495**	0.445	0.506	0.408	0.482	0.493	0.506
	Slight	0.925	**0.963**	0.756	0.627	0.350	**0.963**	0.749	0.758	0.340
FPR	Fatal	0.027	**0.003**	0.148	0.225	0.312	0.004	0.162	0.148	0.312
	Serious	0.051	0.045	0.111	0.151	0.304	**0.038**	0.102	0.107	0.304
	Slight	0.847	0.322	0.252	0.160	**0.110**	0.322	0.247	0.251	0.115
Precision	Fatal	0.092	**0.992**	0.601	0.652	0.516	**0.992**	0.587	0.603	0.514
	Serious	0.297	0.779	0.743	0.602	0.471	**0.805**	0.752	0.746	0.451
	Slight	**0.824**	0.652	0.657	0.662	0.608	0.653	0.661	0.659	0.605
F-Measure	Fatal	0.118	0.906	0.687	0.742	0.593	**0.918**	0.686	0.691	0.591
	Serious	0.153	0.542	**0.594**	0.512	0.488	0.542	0.588	**0.594**	0.482
	Slight	**0.872**	0.778	0.703	0.644	0.444	0.778	0.702	0.705	0.441
AUC	Average	0.619	0.935	0.823	0.818	**0.961**	0.935	0.825	0.825	0.706
MCC	Average	0.103	0.665	0.490	0.465	0.281	**0.674**	0.489	0.493	0.281

TABLE XI
EVALUATION RESULTS FOR THE PROPOSED PREDICTIVE MODELS WITH THE DT CLASSIFIER -DATASET 1

Data Splitting			Holdout				10-folds CV			
			Proposed model-oversampling technique X-DT							
Metric	Class	Imbalanced dataset	SMOTE	SVMSMOTE	BLSMOTE	ADASYN	SMOTE	SVMSMOTE	BLSMOTE	ADASYN
TRP (Recall)	Fatal	0.042	0.906	0.937	0.969	0.940	0.904	0.937	**0.968**	0.938
	Serious	0.221	0.655	0.743	0.712	0.712	0.657	**0.747**	0.711	0.709
	Slight	**0.780**	0.652	0.712	0.666	0.642	0.647	0.711	0.671	0.645
FPR	Fatal	**0.022**	0.075	0.031	0.034	0.060	0.075	0.031	0.032	0.060
	Serious	0.199	0.174	0.173	**0.158**	0.168	0.174	0.173	0.157	0.168
	Slight	0.752	0.149	0.163	0.135	**0.130**	0.149	0.160	0.138	0.132
Precision	Fatal	0.034	0.856	0.896	0.934	0.882	0.858	0.893	**0.938**	0.883
	Serious	0.185	0.656	0.735	0.690	0.693	0.652	**0.736**	0.692	0.693
	Slight	**0.818**	0.690	0.739	0.714	0.709	0.688	0.742	0.714	0.705
F-Measure	Fatal	0.038	0.880	0.916	**0.952**	0.910	0.881	0.915	**0.952**	0.910
	Serious	0.201	0.655	0.739	0.701	0.702	0.654	**0.741**	0.701	0.701
	Slight	**0.798**	0.670	0.725	0.690	0.674	0.667	0.726	0.692	0.674
AUC	Average	0.613	0.952	0.954	0.960	**0.961**	0.802	0.838	0.837	0.822
MCC	Average	0.023	0.606	0.649	0.673	0.645	0.604	0.650	**0.675**	0.643

TABLE XII
EVALUATION RESULTS FOR THE PROPOSED PREDICTIVE MODELS WITH THE ET CLASSIFIER -DATASET 1

Data Splitting			Holdout				10-folds CV			
			Proposed model-oversampling technique X-ET							
Metric	Class	Imbalanced dataset	SMOTE	SVMSMOTE	BLSMOTE	ADASYN	SMOTE	SVMSMOTE	BLSMOTE	ADASYN
TRP (Recall)	Fatal	0.010	0.977	0.977	0.986	0.984	0.976	0.978	**0.987**	**0.987**
	Serious	0.027	0.789	0.781	0.827	**0.835**	0.787	0.781	0.826	0.834
	Slight	**0.983**	0.774	0.868	0.794	0.774	0.77	0.871	0.787	0.774
FPR	Fatal	**0.001**	0.034	0.012	0.012	0.023	0.034	0.011	0.014	0.021
	Serious	**0.017**	0.104	0.079	0.100	0.107	0.103	0.076	0.103	0.11
	Slight	0.974	0.092	0.140	0.085	**0.077**	0.093	0.143	0.085	0.078
Precision	Fatal	0.001	0.931	0.957	**0.975**	0.953	0.934	0.960	0.973	0.956
	Serious	0.246	0.789	0.865	0.804	0.806	0.791	**0.870**	0.799	0.807
	Slight	0.814	0.812	0.799	0.826	**0.831**	0.810	0.798	0.826	0.829
F-Measure	Fatal	0. 002	0.954	0.967	**0.981**	0.968	0.955	0.969	0.980	0.971
	Serious	0.048	0.789	0.821	0.815	0.820	0.790	**0.823**	0.813	0.821
	Slight	**0.890**	0.793	0.831	0.810	0.801	0.793	0.833	0.806	0.801
AUC	Average	0.613	0.952	0.954	0.960	0.961	0.953	0.954	0.961	**0.962**
MCC	Average	0.025	0.770	0.781	**0.803**	0.795	0.770	0.783	0.800	0.796

TABLE XIII
EVALUATION RESULTS FOR THE PROPOSED PREDICTIVE MODELS WITH THE RF CLASSIFIER -DATASET-2

Data Splitting		Holdout				10-folds CV			
		Proposed model-oversampling technique X-RF							
Metric	Class	SMOTE	SVMSMOTE	BLSMOTE	ADASYN	SMOTE	SVMSMOTE	BLSMOTE	ADASYN
TRP (Recall)	Fatal	**0.986**	0.959	0.973	0.968	0.984	0.967	0.976	0.971
	Serious	0.821	0.781	0.722	0.798	**0.857**	0.764	0.753	0.798
	Slight	0.757	**0.816**	0.728	0.711	0.760	**0.816**	0.733	0.715
FPR	Fatal	0.023	0.035	0.053	0.032	**0.019**	0.031	0.048	0.029
	Serious	0.119	**0.097**	0.124	0.138	0.122	0.100	0.121	0.115
	Slight	0.087	0.151	0.114	0.091	**0.071**	0.143	0.100	0.080
Precision	Fatal	**0.991**	0.887	0.899	0.939	0.958	0.898	0.911	0.941
	Serious	**0.885**	0.833	0.744	0.733	0.778	0.831	0.757	0.736
	Slight	**0.898**	0.772	0.766	0.803	0.859	0.786	0.785	0.805
F-Measure	Fatal	**0.992**	0.922	0.935	0.954	0.971	0.931	0.943	0.956
	Serious	**0.873**	0.788	0.733	0.764	0.816	0.796	0.755	0.766
	Slight	**0.906**	0.794	0.747	0.754	0.807	0.801	0.758	0.758
AUC	Average	**0.978**	0.933	0.927	0.934	0.950	0.938	0.934	0.940
MCC	Average	**0.867**	0.741	0.707	0.741	0.795	0.752	0.730	0.744

TABLE XIV
EVALUATION RESULTS FOR THE PROPOSED PREDICTIVE MODELS WITH THE KNN CLASSIFIER -DATASET 2

Data Splitting		Holdout				10-folds CV			
		Proposed model-oversampling technique X-KNN							
Metric	Class	SMOTE	SVMSMOTE	BLSMOTE	ADASYN	SMOTE	SVMSMOTE	BLSMOTE	ADASYN
TRP (Recall)	Fatal	**0.987**	0.969	0.974	0.965	0.985	0.970	0.979	0.967
	Serious	0.869	0.785	0.767	0.811	**0.893**	0.800	0.783	0.795
	Slight	0.590	0.709	0.620	0.470	0.604	**0.712**	0.622	0.570
FPR	Fatal	0.034	0.052	0.074	0.087	**0.027**	0.045	0.066	0.066
	Serious	0.202	**0.152**	0.166	0.237	0.201	0.154	0.165	0.195
	Slight	0.056	0.116	0.083	0.052	**0.049**	0.113	0.077	0.072
Precision	Fatal	0.928	0.842	0.864	0.850	**0.940**	0.857	0.881	0.882
	Serious	0.677	0.769	0.699	0.618	0.690	**0.770**	0.704	0.660
	Slight	0.856	0.794	0.794	0.824	**0.876**	0.802	0.801	0.803
F-Measure	Fatal	0.956	0.901	0.916	0.904	**0.962**	0.910	0.928	0.923
	Serious	0.761	0.777	0.731	0.702	0.778	**0.785**	0.741	0.721
	Slight	0.699	0.749	0.696	0.599	0.715	**0.754**	0.700	0.667
AUC	Average	0.911	0.923	0.909	0.934	**0.929**	0.927	0.917	0.908
MCC	Average	0.721	0.704	0.678	0.638	**0.740**	0.714	0.694	0.673

TABLE XV
EVALUATION RESULTS FOR THE PROPOSED PREDICTIVE MODELS WITH THE NB CLASSIFIER -DATASET 2

Data Splitting		Holdout				10-folds CV			
		Proposed model-oversampling technique X-NB							
Metric	Class	SMOTE	SVMSMOTE	BLSMOTE	ADASYN	SMOTE	SVMSMOTE	BLSMOTE	ADASYN
TRP (Recall)	Fatal	0.827	0.730	0.792	**0.974**	0.832	0.724	0.787	**0.974**
	Serious	0.587	0.557	0.397	0.043	**0.618**	0.550	0.404	0.043
	Slight	0.624	0.759	0.681	0.192	0.605	**0.783**	0.670	0.192
FPR	Fatal	**0.085**	0.114	0.188	0.812	0.087	0.102	0.184	0.822
	Serious	0.224	0.140	0.177	**0.019**	0.235	0.136	0.182	**0.019**
	Slight	0.183	0.249	0.201	**0.053**	0.165	0.261	0.203	0.054
Precision	Fatal	**0.814**	0.645	0.671	0.377	0.807	0.664	0.682	0.377
	Serious	0.561	0.720	0.529	0.522	0.568	**0.722**	0.526	0.522
	Slight	0.660	0.657	0.634	0.650	**0.677**	0.659	0.623	0.650
F-Measure	Fatal	**0.821**	0.685	0.727	0.544	0.819	0.693	0.730	0.544
	Serious	0.574	**0.628**	0.453	0.080	0.592	0.624	0.457	0.080
	Slight	0.642	0.705	0.657	0.296	0.639	**0.716**	0.646	0.296
AUC	Average	0.845	0.828	0.809	0.688	**0.850**	0.827	0.807	0.687
MCC	Average	0.509	0.497	0.430	0.181	**0.514**	0.504	0.428	0.181

TABLE XVI
EVALUATION RESULTS FOR THE PROPOSED PREDICTIVE MODELS WITH THE DT CLASSIFIER -DATASET 2

Data Splitting		Holdout				10-folds CV			
		Proposed model-oversampling technique X-DT							
Metric	Class	SMOTE	SVMSMOTE	BLSMOTE	ADASYN	SMOTE	SVMSMOTE	BLSMOTE	ADASYN
TRP (Recall)	Fatal	0.962	0.969	0.975	0.962	0.969	0.975	0.976	**0.976**
	Serious	0.751	0.790	0.773	0.750	0.787	0.772	**0.795**	0.786
	Slight	0.630	**0.717**	0.643	0.626	0.715	0.642	0.636	0.625
FPR	Fatal	0.059	**0.030**	0.049	0.060	**0.030**	0.052	0.041	0.044
	Serious	0.170	0.167	0.167	0.172	0.168	**0.166**	0.170	0.173
	Slight	0.102	0.123	0.090	0.100	0.126	0.089	**0.084**	0.089
Precision	Fatal	0.887	0.900	0.907	0.889	0.900	0.903	**0.924**	0.920
	Serious	0.684	**0.750**	0.692	0.679	0.748	0.693	0.689	0.683
	Slight	0.766	0.792	0.787	0.763	0.788	0.789	**0.796**	0.784
F-Measure	Fatal	0.924	0.933	0.940	0.924	0.934	0.938	0.949	**0.947**
	Serious	0.716	**0.769**	0.730	0.713	0.767	0.731	0.738	0.731
	Slight	0.691	**0.753**	0.708	0.688	0.750	0.692	0.707	0.696
AUC	Average	0.921	0.932	0.928	0.866	0.884	0.877	**0.934**	0.882
MCC	Average	0.673	0.693	0.697	0.670	0.690	0.696	0.707	**0.697**

TABLE XVII
EVALUATION RESULTS FOR THE PROPOSED PREDICTIVE MODELS WITH THE ET CLASSIFIER -DATASET 2

Data Splitting		Holdout				10-folds CV			
		Proposed model-oversampling technique X-ET							
Metric	Class	SMOTE	SVMSMOTE	BLSMOTE	ADASYN	SMOTE	SVMSMOTE	BLSMOTE	ADASYN
TRP (Recall)	Fatal	0.974	0.969	**0.978**	**0.978**	0.974	0.970	**0.978**	**0.978**
	Serious	0.787	0.801	0.794	0.814	0.780	0.797	0.784	**0.810**
	Slight	0.712	0.810	0.726	0.717	0.703	**0.813**	0.728	0.712
FPR	Fatal	0.034	**0.012**	**0.012**	0.032	0.045	0.022	0.040	0.033
	Serious	0.104	**0.079**	0.100	0.133	0.139	0.112	0.128	0.135
	Slight	0.092	0.140	0.085	**0.079**	0.090	0.122	0.088	0.081
Precision	Fatal	0.916	0.923	0.923	**0.939**	0.916	0.924	0.924	0.938
	Serious	0.741	0.818	0.749	0.744	0.732	**0.819**	0.749	0.740
	Slight	0.810	0.816	0.821	**0.823**	0.803	0.814	0.811	0.820
F-Measure	Fatal	0.944	0.945	0.950	**0.958**	0.944	0.946	0.950	**0.958**
	Serious	0.763	**0.810**	0.771	0.777	0.755	0.808	0.766	0.773
	Slight	0.758	0.813	0.771	0.767	0.750	**0.814**	0.768	0.762
AUC	Average	0.921	0.932	0.928	0.934	0.921	0.932	0.928	**0.935**
MCC	Average	0.736	0.755	0.749	**0.756**	0.729	0.754	0.745	0.752

After balancing the dataset-2 using different oversampling techniques, SMOTE, SVMSMOTE, BLSMOTE, and ADASYN, they are analyzed using different classifiers, RF, KNN, NB, DT, and ET, for the prediction of accident severity. The proposed model-SMOTE-RF and proposed model-SVMSMOTE-RF outperformed the others according to the comparative metrics as shown in Table XIII. By observing the results in Table XIV, the same conclusion is obtained. The Holdout data splitting method is suitable with the proposed model-SMOTE-RF, whereas the 10 folds CV is suitable with the proposed model-SMOTE-KNN. Table XV shows the evaluation results for the proposed predictive models with the NB classifier. These models prove that SMOTE, SVMSMOTE, and ADASYN cause a significant enhancement than BLSMOTE. The results in Table XVI indicate that all models enhance predicting each class, and there is no superiority based on all used metrics between them, whereas, in Table XVII, the proposed model-SMOTE-ET has low performance than others. In all the cases, the best performance of the classifier is RF, followed by KNN and ET, and then by NB and DT.

For companies' real application of such models, this is reflected in a more accurate prediction about fatal accidents or those causing injury and a certain number of false alarms (the FP instances) that are mistakenly classified as severe.

The output of the classifiers is the accident severity class (Fatal, Serious, or Slight). In [14], the two minor classes (Fatal and Serious) are consolidated as one class. We named the "Fatal and Serious " class as "A," whereas the "Slight" class as "B." In [14], the dataset splitting approach is the holdout (70%, 30%). We consolidated the classes before applying the proposed predictive model as presented in Tables XVIII and XIX for comparison purposes. They show the performance of models in predicting each class. In the imbalanced dataset, predicting the minor class (A) is very weak. Whereas in the case of RUMS-based classifiers, the result of the classifiers decreases for predicting class (B) and increases for predicting class (A). In the case of the proposed predictive model, the performance of the models significantly increases for predicting both classes.

TABLE XVIII
COMPARISON FOR THE PROPOSED PREDICTIVE MODELS WITH THE RF CLASSIFIER -DATASET 2

				Proposed model-oversampling technique X-RF			
Metric	Class	Imbalanced dataset-RF[14]	RUMC-RF[14]	SMOTE	SVMSMOTE	BLSMOTE	ADASYN
TRP (Recall)	A	0.058	0.568	0.865	0.780	**0.866**	0.823
	B	**0.962**	0.560	0.819	0.803	0.808	0.747
FPR	A	**0.038**	0.440	0.181	0.197	0.192	0.253
	B	0.942	0.432	0.135	0.220	**0.134**	0.177
Precision	A	0.203	0.176	**0.823**	0.800	0.814	0.763
	B	0.860	**0.887**	0.862	0.784	0.862	0.811
F-Measure	A	0.090	0.269	0.843	0.789	**0.839**	0.792
	B	**0.909**	0.687	0.840	0.793	0.834	0.778

TABLE XVIII
COMPARISON FOR THE PROPOSED PREDICTIVE MODELS WITH THE KNN CLASSIFIER -DATASET 2

				Proposed model-oversampling technique X-KNN			
Metric	Class	Imbalanced dataset-KNN[14]	RUMC-KNN[14]	SMOTE	SVMSMOTE	BLSMOTE	ADASYN
TRP (Recall)	A	0.183	0.572	**0.907**	0.822	0.904	0.869
	B	**0.885**	0.470	0.672	0.710	0.661	0.520
FPR	A	**0.115**	0.530	0.328	0.290	0.339	0.480
	B	0.817	0.428	**0.093**	0.178	0.096	0.132
Precision	A	0.209	0.152	0.729	**0.741**	0.721	0.641
	B	0.867	0.869	**0.881**	0.799	0.876	0.800
F-Measure	A	0.195	0.240	**0.808**	0.779	0.802	0.738
	B	**0.876**	0.610	0.763	0.752	0.754	0.630

The results indicated that reasonable prediction is obtained from applying the proposed predictive model-BLSMOTE-RF, and the proposed predictive model-SMOTE-KNN. They also reveal that when applying the proposed predictive model-RF or the proposed model-KNN classifiers on a different dataset, there is a significant enhancement in performance for predicting the minority classes. Indeed, the proposed predictive model can predict satisfying both class (A) and class (B).

V. CONCLUSION

With realizing the importance of traffic safety and sustainable transportation in our lives, we demonstrated a new proposed predictive model for traffic accident severity prediction in this paper. Different models (RF, KNN, NB, DT, ET) classifiers have been employed after solving the class imbalance problem through SMOTE, SVMSMOTE, BLSMOTE, and ADASYN. Two real-world datasets have been utilized for training and testing these models. Several evaluation metrics have been employed to confirm the proposed predictive model that predicts minor and major classes. According to the results of the proposed predictive model-SMOTE-RF (10-folds CV) using dataset-1, the accuracy is 92.80 %, Kappa statistics is 0.8909, and the training time is 69.44 seconds. Using dataset-2, the accuracy is 86.04%, Kappa statistics is 0.7904, and the training time is 3.65 seconds. In comparison, the proposed predictive model-KNN and proposed predictive model-NB are fast compared to others proposed models. The superior performance of our proposed predictive model revealed that it could be used as a reliable and robust tool that improves traffic safety management.

REFERENCES

[1] World Health Organization, “Save lives: a road safety technical package,” 2017

[2] X. MENG, L. ZHENG, and G. QIN, “Traffic accidents prediction and prominent influencing factors analysis based on fuzzy logic,” *Journal of Transportation Systems Engineering and Information Technology*, vol. 2, pp1-15, 2009

[3] P. T. Savolainen, F. L. Mannering, D. Lord, and M. A. Quddus, “The statistical analysis of highway crash-injury severities: a review and assessment of methodological alternatives,” *Accident Analysis & Prevention*, vol. 43, no. 5, pp1666–1676, 2011

[4] K. Li, D. Qian, S. Huang, and X. Liang, “Analysis of traffic accidents on highways using latent class clustering,” *in Proceedings of the 16th COTA International Conference of Transportation Professionals, Shanghai, China,* 2016, pp1800–1810

[5] B. Dadashova, B. A. Ramírez, J. M. McWilliams, and F. A. Izquierdo, “The identification of patterns of interurban road accident frequency and severity using road geometry and traffic indicators,” *Transportation Research Procedia*, vol. 14, pp4122-4129, 2016

[6] Z. Yan, X. Lu, and W. Hu, “Analysis of factors affecting traffic accident severity based on heteroskedasticity ordinal Logit,” in *Proceedings of the 6th International Conference on Transportation Engineering, Chengdu, China,* 2020, pp422–435

[7] S. Cateni, V. Colla, and M. Vannucci, “A method for resampling imbalanced datasets in binary classification tasks for real-world problems,” *Neurocomputing*, vol. 135, pp32-41, 2014

[8] N. Japkowicz and S. Stephen, “The class imbalance problem: A systematic study,” *Intelligent Data Analysis*, vol. 6, no. 5, pp429–449, 2002

[9] H. He and E. A. Garcia, “Learning from imbalanced data,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp1263–1284, 2009.

[10] R. C. Prati, G. E. Batista, and M. C. Monard, “Data mining with imbalanced class distributions: Concepts and methods,” in *Proceedings of the 4th Indian International Conference on Artificial Intelligence, Karnataka, India, 2009*, pp359-376

[11] H. Guo and H. L. Viktor, “Learning from imbalanced data sets with boosting and data generation: the databoost-im approach,” *ACM Sigkdd Explorations Newsletter*, vol. 6, no. 1, pp30–39, 2004.

[12] N. V Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: synthetic minority over-sampling technique,” *Journal of Artificial Intelligence Research*, vol. 16, pp321–357, 2002.

[13] H. Han, W.Y. Wang, and B.H. Mao, “Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning,” in *International Conference on Intelligent Computing, Hefei, China,* 2005, pp878-887

[14] N. Fiorentini and M. Losa, “Handling imbalanced data in road crash severity prediction by machine learning algorithms,” *Infrastructures*, vol. 5, no. 7, 2020

[15] S. Haynes, P. C. Estin, S. Lazarevski, M. Soosay, and A.-L. Kor, “Data Analytics: Factors of Traffic Accidents in the UK,” in

Proceedings of the 10th International Conference on Dependable Systems, Services, and Technologies, Leeds, United Kingdom, 2019, pp120-126

[16] F. Lécué, S. Tallevi-Diotallevi, J. Hayes, R. Tucker, V. Bicer, M. Sbodio, and P. Tommasi, "Smart traffic analytics in the semantic web with STAR-CITY: Scenarios, system, and lessons learned in Dublin City," *Journal of Web Semantics*, vol. 27, pp 26-33, 2014

[17] T. Beshah, D. Ejigu, A. Abraham, V. Snášel, and P. Krömer, "Knowledge discovery from road traffic accident data in Ethiopia: data quality, ensembling and trend analysis for improving road safety," *Neural Network World*, vol. 22, no. 3, pp215-244, 2012

[18] T. Beshah, A. A. Dejene Ejigu, V. Snasel, and P. Kromer, "Mining pattern from road accident data: role of road user's behaviour and implications for improving road safety," *International Journal of Tomography and Simulation*, vol. 22, no. 1, pp73-86, 2013

[19] T. Beshah, D. Ejigu, A. Abraham, V. Snasel, and P. Kromer, "Pattern recognition and knowledge discovery from road traffic accident data in Ethiopia: implications for improving road safety," in *Proceedings of the 2011 World Congress on Information and Communication Technologies*, *Mumbai, India*, 2011, pp1241-1246

[20] T. Beshah and S. Hill, "Mining road traffic accident data to improve safety: role of road-related factors on accident severity in Ethiopia," in *Proceedings of the AAAI Spring Symposium: Artificial Intelligence for Development, Stanford, California, USA*, 2010, vol. 24, pp1173-1181

[21] S. Krishnaveni and M. Hemalatha, "A perspective analysis of traffic accident using data mining techniques," *International Journal of Computer Applications*, vol. 23, no. 7, pp40-48, 2011

[22] 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*, CA, USA, 2014, pp322-329

[23] C. S. Perone, "Injury risk prediction for traffic accidents in Porto Alegre/RS, Brazil," *arXiv preprint arXiv:1502.00245*, 2015

[24] C. Chen, G. Zhang, Z. Qian, R. A. Tarefder, and Z. Tian, "Investigating driver injury severity patterns in rollover crashes using support vector machine models," *Accident Analysis & Prevention*, vol. 90, pp128-139, 2016

[25] A. A. Qasem and J. D. Esraa, "Data mining methods for traffic accident severity prediction," *International Journal of Neural Networks and Advanced Applications*, vol. 5, pp1-12, 2018

[26] S. Ramya, S. K. Reshma, V. D. Manogana, and Y. S. Saroja, "Accident severity prediction using data mining methods," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 5, no. 2, 2019

[27] L. Wahab and H. Jiang, "Severity prediction of motorcycle crashes with machine learning methods," *International Journal of Crashworthiness*, vol. 25, no. 5, pp485-492, 2020

[28] V. López, A. Fernández, S. García, V. Palade, and F. Herrera, "An insight into classification with imbalanced data: empirical results and current trends on using data intrinsic characteristics," *Information Sciences*, vol. 250, pp113-141, 2013

[29] H. M. Nguyen, E. W. Cooper, and K. Kamei, "Borderline over-sampling for imbalanced data classification," *Journal of Knowledge Engineering and Soft Data Paradigms*, vol. 3, pp4-21, 2011

[30] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: adaptive synthetic sampling approach for imbalanced learning," in 2008 *IEEE International Joint Conference on Neural Networks* (IEEE World Congress on Computational Intelligence), 2008, pp1322-1328

[31] R. Longadge and S. Dongre, "Class imbalance problem in data mining review," *arXiv preprint arXiv:1305.1707*, 2013

[32] B. Krawczyk, M. Woźniak, and G. Schaefer, "Cost-sensitive decision tree ensembles for effective imbalanced classification," *Applied Soft Computing*, vol. 14, pp554-562, 2014

[33] H. Kaur, H. S. Pannu, and A. K. Malhi, "A systematic review on imbalanced data challenges in machine learning: applications and solutions," *ACM Computing Surveys (CSUR)*, vol. 52, no. 4, pp1-36, 2019

[34] M. Pérez-Ortiz, S. Jiménez-Fernández, P. A. Gutiérrez, E. Alexandre, C. Hervás-Martínez, and S. Salcedo-Sanz, "A review of classification problems and algorithms in renewable energy applications," *Energies*, vol. 9, no. 8, 2016

[35] L. Breiman, "Random Forests" *Machine Learning*, vol. 45, pp5-32, 2001

[36] N. Jiang, F. Fu, H. Zuo, X. Zheng, and Q. Zheng, "A Municipal PM2.5 Forecasting Method Based on Random Forest and WRF Model, " *Engineering Letters*, vol. 28, no. 2, pp312-321, 2020

[37] N. Dogru and A. Subasi, "Traffic accident detection using random forest classifier," in *Proceedings of the 15th Learning and Technology Conference, Jeddah, Saudi Arabia*, 2018, pp40–45

[38] M. Asif, M. M. Nishat, F. Faisal, R. R. Dip, M. H. Udoy, M. Shikder, and R. Ahsan, "Performance evaluation and comparative analysis of different machine learning algorithms in predicting cardiovascular disease, " *Engineering Letters*, vol. 29, no. 2, pp731-741, 2021

[39] T. Georgieva-Trifonova, M. Stefanova, and S. Kalchev, "Customer feedback text analysis for online stores reviews in Bulgarian, " *IAENG International Journal of Computer Science*, vol. 45, no. 4, pp560-568, 2018

[40] R.Lior , and O. Z. Maimon, "Data mining with decision trees: theory and applications, *" World Scientific*, vol. 69. 2007

[41] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," Machine Learning, vol. 63, no. 1, pp3-42, 2006

[42] F. J. Valverde-Albacete, J. Carrillo-de-Albornoz, and C. Peláez-Moreno, "A proposal for new evaluation metrics and result visualization technique for sentiment analysis tasks," in *International Conference of the Cross-Language Evaluation Forum for European Languages, Berlin, Germany*, 2013, pp41-52

[43] B. J. M. Abma, "Evaluation of requirements management tools with support for traceability-based change impact analysis," *Master's thesis, University of Twente, Enschede*, 2009

[44] M. Batista, B. Oliveira, P. Chaves, J. C. Ferreira, and T. Brandao, "Improved real-time wildfire detection using a surveillance system, " in *Proceedings of the World Congress on Engineering*, London, U.K., 2019

[45] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, pp159-174, 1977

[46] M. Grandin, E. Bagli, and G.Visani, "Metrics for Multi-Class Classification: an Overview, " *arXiv preprint arXiv: 2008.05756*, 2020

[47] The United Kingdom Open Data Portal, https://data.gov.uk (accessed on March 1, 2021)