J. Electrical Systems 20-6s (2024): 2733-2746

[1]* **Dr. Rudresh Deepak Shirwaikar**

[2]**Parthiv K P**

[2]**Akarsh H Simha**

[1]**Ashish Narvekar**

[3]**Tumbalam Gooty Pradeep**

Predicting Accident Severity using Machine Learning

Abstract: - Each year, millions of people are killed in automobile accidents. Predicting the severity of an occurrence allows local authorities to respond quickly and save many lives. Information and data on traffic accidents made available by public organizations can be used to categories these incidents based on their nature and severity, and then attempt to construct predictive models that can be further investigated to identify fatal accident risk factors. The study provides ways to develop a system for determining the severity of accidents. To evaluate the severity of the collision, we use a variety of weather-related characteristics such as temperature, humidity, visibility, and pressure, as well as several other circumstances, such as the presence of a traffic light or a junction. The Select-K-Best features selection algorithms were used to choose the best characteristics from a list of 47. The accuracy of both balanced and unbalanced data was measured, and the balanced data was chosen for further analysis. The balanced dataset is then trained, analyzed, and compared with a number of machine learning approaches, including the Random Forest Classifier (RFC), K closest neighbour (KNN) classifier, and Naive Bayes classifier. The RFC classifier outperformed with an AUC of 95%, whereas the naïve bayes underperformed. Furthermore, the accuracy is improved by using all of the aforementioned methods as base learners for stacking ensemble models with logistic regression as meta learner. This stacking ensemble strategy outperforms RFC with improved precision and accuracy, resulting in an AUC of 96.92%. The results revealed that the model has a 96.92% AUC in predicting accident severity. The work can be extended to investigate the complex correlations between a few key parameters and accident severity.

Keywords: Machine learning, Accident Severity, Unbalanced Datasets, Feature Selection

I. Introduction

According to a WHO, vehicle traffic accidents kill over 1.3 million people each year [1]. An additional 20 to 50 million people suffer non-fatal injuries, and many of them go on to become disabled. Despite the fact that some nations have over 60% of the world's vehicles, 93% of all road deaths take place in low- and middle-income nations. The layout of a road can have a big impact on how safe it is. Road traffic accidents cost the majority of countries 3 percent of their GDP. For victims, their families, and entire nations, traffic accidents cause substantial economic harm. These expenditures are the consequence of the cost of medicine, lost wages for those who pass away or become disabled due to injuries, and caregiving costs for loved ones. Since they cause many injuries and fatalities, road accidents have a substantial impact on society.

In recent years, research has increasingly focused on determining the significance of the seriousness of driver injuries caused by traffic accidents. Accident analysis should be based on precise and comprehensive accident records. Accidents happen and they are frequently unpredictable and unavoidable, but helping the victims and providing support when it is needed can save many lives Our study is concentrated on creating a machine learning model that can foresee the severity of incidents.

[1] *Dr Rudresh Shirwaikar, Professor and Head, Dept of Computer Engineering, Agnel Institute of Technology and Design, Goa.

[2] Parthiv K, Dept of Information Science and Engineering, M S Ramaiah Institute of Technology, Bangalore.

[2] Akarsh H Simha, Dept of Information Science and Engineering, M S Ramaiah Institute of Technology, Bangalore.

[1] Ashish Narvekar, Dept of Computer Engineering, Agnel Institute of Technology and Design, Goa

[3] Tumbalam Gooty Pradeep, BMSIT, Bangalore

Due to the traffic officials' and paramedics' tardy arrival, several lives are lost. However, if the seriousness of accidents is known ahead of time in such locations or under such circumstances, prompt assistance can be given. The goal is to inform concerned officials and other local authorities of an accident's significance in advance so that they can take the appropriate measures, such as stationing emergency vehicles like ambulances, police cars, and fire engines for rapid assistance. Recognizing objects and classifying them is a common job for machine learning algorithms. This classification a supervised technique [2] allows us to separate massive volumes of data into discrete values or a set of specified output label classes [3].

The study focuses on creating a multiclassification model utilizing machine learning technique. K-nearest Neighbors, Random Forest classifier [4], Naive Bayes, Random Forest [5], and lastly Stacking ensemble approach [6] are among the techniques proposed for the study. The SMOTE algorithm [7] is employed to deal with the data set's imbalance. After addressing the imbalance in the data, the study aims to improve the model's performance and forecast the severity of accidents. This will allow for the efficient use of time and resources by stationing emergency vehicles in areas and situations where the severity rate of accidents is high (and fatalities are likely to occur).

II. RELATED WORK

The section highlights the important work caried out in the area of accident severity prediction using machine learning approach. It also describes the features that affects the severity of accidents. Furthermore, it compares various methods used to handle unbalance in data.

The study carried out by Cuenca et al. [8] identified the various important factors affecting accident severity and had used these features to fit the machine learning models namely Naive Bayes, Boosting Trees, and Deep Architectures to predict the severity of the accident on the dataset that was collected from the Spanish traffic agency from 2011 to 2015. The dataset collected for the study was highly imbalanced so subsampling technique was employed to obtain a lesser amount of data that belonged to the majority class without affecting the elements of the minority class. Various machine learning and statistical models were evaluated by Zhang et al. [9] for their ability to predict crash injury severity. The authors were able to determine the significance of various variables/features on crash severity through the study of sensitivity and compare the carriable relevance using various approaches. Problems with multiclass categorization were the focus of the study. The models used had the goal of categorizing the accidents into 5 distinct levels of severity. Level 1 severity denotes no injury, Level 2 severity suggests a potential or invisible injury, Level 3 denotes an injury that is not incapacitating, Level 4 denotes an injury that is incapacitating, and Level 5 denotes a fatal injury. The 10 most significant traits that were significantly associated with crash severity were identified using the feature selection method. Multinomial logistic regression (MNL) and One factor analysis served as the statistical models used in this study. KNN, Decision Trees, RF, and SVM were among the machine learning models used. The RF achieved the greatest testing accuracy of 53.9 percent, while the Multinomial Logistic Regression model, a statistical model, achieved the testing accuracy of 50.9 percent. The outcomes of the various models' predictions indicate that machine learning models outperform statistical models.

Labib et al. [10] used four machine learning approaches to assess traffic incidents and determine the severity of the accidents. A total of 15 features were chosen from the initial 34 accessible features using a variety of feature selection techniques, which were based on feature importance. The accidents were divided into four groups: motor vehicle collisions, fatalities, grievous incidents, and simple injuries. Class 1 (Fatal) samples made up the majority of the dataset used for the study, which made it unbalanced. Due to this class, which integrated the severity of grievous, simple injury, and motor vehicle collisions, new projections were generated on the revised category targets. Decision Tree, Naive Bayes, KNN and AdaBoost were the supervised machine learning techniques employed. The AdaBoost algorithm performs best in both multiclass and binary classification. Recurrent neural networks were used in a study by Sameen et al. in [11] to anticipate the accident's severity. Property damage, obvious injury, and disabling injury classes were used to categories the accident severity. By adding Gaussian noise to the training data, using ReLU as the activation function in the hidden layers, and ultimately utilizing the dropout approach, the overfitting issue in the RNN model was resolved. The highest accuracy of 74.67 percent was obtained by using Adam as the optimizer.

Using genetic algorithms and neural networks, Li et al. [12] found key elements that contribute to the severity of accidents. The multi-objective optimization method was effectively applied in conjunction with the Neural Network based Non-Dominated Sorting Algorithm (NSGA-II) architecture. According to the study, the crucial characteristics that determine an accident's severity are its road surface, as opposed to lighting conditions, the time of year, and day of the week, which had no discernible effect.

Research conducted by Joaquin et al. in [13] was responsible for extracting a total of 70 relevant rules using several choice tree algorithms. This was done since data rules are dependent on the decision tree structure, therefore selecting the features based on a single decision tree was not viable. To improve various environmental and infrastructure features, such as safety barriers, shoulders, visibility, and lighting, to primarily minimize the frequency of accidents and to lessen the severity of the accidents, every feasible set of data rules were acquired from each of the potential trees, and the resulting beneficial rules were transmitted to the involved authorities. In order to estimate the accident severity, Yan et al [14] made use of accident-related data from Pennsylvania's Montgomery County. Three categories—slight, serious, and fatal were used to categories the accident's severity. For achieving the initial prediction results, conventional algorithms such as ANN, KNN, SVM, and RF were utilized. In order to fit the most important features, the study further used the RF as the base algorithm. The RF was then optimized using the Bayesian optimization technique. Mamlook et al. [15] identified a set of influential features and built classification models using a variety of supervised machine learning techniques, the algorithms that were used in the study were AdaBoost, Logistic Regression, Naive Bayes and Random Forest. Synthetic Minority Oversampling Technique (SMOTE) algorithm was used to handle the imbalance that existed in the original dataset. The highest accuracy of 75.5% and a precision score of 0.82 was achieved by the RF for the binary classification task. Comparing the hybrid model to conventional machine learning methods, the hybrid model was able to offer superior precision.

The dataset gathered from Knox County in Tennessee was in literature [16]. The suggested model consists of an unsupervised feature learning module to identify functional networks between the explanatory factors and the feature representations, and a supervised fine-tuning module for traffic crash prediction. To address the unobserved diverseness challenges in the traffic crash prediction, a multivariate negative binomial (MVNB) model was incorporated as a regression layer into the supervised fine-tuning module. The model divided the accident into three groups: significant injuries, minor injuries, and no injuries. The results show that the proposed algorithm outperformed the SVM model and the deep learning architecture. The proposed algorithm was able to reach the best accuracy of 96.63 percent.

Chawla et al. [17] suggested creating new data points rather than duplicated data points in order to oversample the samples of the minority class. The KNN algorithm was used in the strategy to generate synthetic data points. According to the study's execution, the value of k was set to 5. Several datasets with imbalance ranging from low to high levels were used to test the proposed technique, however the accuracy of KNN on the dataset on which the SMOTE technique was tested revealed better prediction outcomes.

One of the key elements of traffic accidents is spatiotemporal correlation, which Ren et al. [18] originally presented and then used to build a highly accurate deep architectures for analyzing the likelihood of a traffic collision. To avoid overfitting, a dropout layer with a rate of 0.5 between the connected layers of the deep learning model was built using a unique strategy based on LSTM architecture with ReLU activation function. Three of the four LSTM layers in the model were fully linked. Lasso, SVM, RFR, and other baseline models were used to compare the performance of the unique technique. The accuracy and precision of the prediction were assessed using three metrics: mean absolute error (MAE), mean squared error (MSE), and root mean square error (RMSE). MAE, MSE, and RMSE values for the novel approach were 0.014, 0.001, and 0.34 respectively, outperforming the basic models.

On the basis of data collected for the traffic incidents that happened in Setubal, Portugal, Santos et al. [19] determined the causes of road accidents. For predicting the severity of road accidents, a variety of machine learning algorithms including RF, DT, decision trees, logistic regression, and naive bayes were used. Unsupervised machine learning techniques such as DBSCAN and hierarchical clustering were also used. The choice of features was based on the mean decrease values (MDA), which show how accurate the model becomes when a certain variable or feature is removed. With a 73 percent accuracy rate and an AUC Score of 0.68, the study's findings showed that the RF serves the best for classifying data.

Theofilatos et al. [20] examined the severity of road accidents by a macroscopic analysis utilizing logistic regression analysis. Using Greek Road accident data from 2008, two models were created to predict the severity of accidents that occurred in and outside of metropolitan areas. Due to inexperience and the fact that younger drivers took more risks when driving, it was determined from the feature importance that the age of the driver was a significant contributing factor for traffic accidents. The investigation also revealed that accidents were more severe at night because of inadequate lighting and low visibility, and that accidents at non-intersections were more severe than accidents at intersections because of faster speeds. Two forms of collisions, sideswipe and rear-end collisions, were seen to be widespread among these two models. There were 4 different types of collisions outside of metropolitan areas and 3 different types inside. The odds ratio showed that rear-end and sideswipe collisions outside of metropolitan areas were more severe.

Statistics on accidents on Taiwan's National Freeway 1 from 2001 to 2002 were acquired for the study by Cheng et al. in [21]. A CART algorithm and a binomial regression (negative) model were built to investigate the empirical association between highway geometry variables and traffic accidents, characteristics of traffic, and environmental factors. It was shown that the CART model had accuracy for both training and testing of 55%, outperforming the binomial regression model in classification. The FP-tree approach was utilized to pick the features, and the features were then used to fit the k-NN model and the Bayesian network to forecast the accident risk. The dataset for the study was the traffic accident data gathered on a section of the I-64 in Norfolk, Virginia, (2005). The highest sensitivity was achieved by the Bayesian network model with NED number 4, and the lowest false alarm rate was 38.16 percent. The performance of the model was rather respectable considering that feature selection using the FP-tree algorithm was a novel approach, and the goal of the study going forward was to enhance the performance of the novel approach.

In order to increase the prediction accuracy on the Wisconsin Breast Cancer Diagnostic (WBCD) data that was provided by the University of Wisconsin, Hyunjin et al. [22] used a stacking ensemble learning approach. Generalized Linear Model (GLM), Gradient Boosting Machine (GBM), Distributed Random Forest (DRF), and Deep Neural Network (DNN) were the base learners used in the study. Each of these models were used independently as a meta learner for stacking ensemble models in order to obtain the highest model performance metrics. The highest average prediction accuracy was attained by the stacking ensemble classifier using the meta learner as the generalized linear model (GLM), which had a 97.37 percent accuracy rating. The accuracy metric of the stacking model was greater than that of the individual models. The summary of literature with objectives, data source, methodology and results are mentioned in Table 1.

Table 1 Summary of literature for use of machine learning for accident severity prediction.

Literature	Objective	Dataset source	Methodology	Results
Cuenca et al. [8]	Traffic accidents classification and injury severity prediction	DGT (Spanish Traffic Agency)	Naive Bayes, Gradient Boosting Trees, Deep Learning with 10 epochs (rectifier, tanh, Maxout, ExpRectifier)	Highest accuracy achieved by Deep Learning (10 epochs -tanh) with accuracy 0.87 and precision 0.85
Zhang et al. [9]	Comparing predictive performance among various machine learning and statistical methods	Florida State traffic department (USA)	DT, RF, KNN, SVM, OP, MNL	ML models performed better than statistical models with highest accuracy achieved by RF at 0.53 for testing.
Labib et al. [10]	Bangladesh based study on acident severity analysis	ARI of BUET (2001-2005)	DT, KNN, Naive Bayes, Ada-Boost	Highest accuracy for both four class and two class classification

	using ML models			achieved by Ada-Boost at 80%
Sameen et al. [11]	Traffic Accident severity prediction using RNN	North-South Expressway accident data (2009-2015)	Recurrent Neural Network (RNN), optimization=SGD, learning rate=0.01, momentum=0.80 and weight decay =0.9	RNN model of 71.77%. accuracy highest among all.
Yan et al. [14]	Traffic Accident Severity prediction using Random Forest	Subset of US-Accidents dataset for Montgomery county (2016-2019)	ANN, KNN, SVM, RF and BO-RF	Random forest had the highest precision, with bayesian optimization RF was able to achieve higher recall, f1 score and AUC score of 0.9625
AlMamlook et al. [15]	Comparing machine algorithms for predicting Traffic accident Severity	Dataset was obtained from the office of Highway Safety Planning	Logistic Regression, RF, Naive Bayes, Ada-Boost	Highest accuracy was achieved by Random Forest classifier and has the highest AUC score
Dong et al. [16]	Improving deep learning models to predict accident severity	Achieved from Tennessee Roadway Information Management System (TRIMS) and PMS	SVM, Deep learning model, Deep learning model with regression layer	Deep learning regression based novel approach (84.21% accuracy) found to be better than deep learning without the regression layer and SVM.

Through the review of the literature, it was determined that researchers did not address the imbalance in the dataset, leading to overfitting in some instances, resulting in average performance metrics, and adding unconscious bias among the classification models. By balancing dataset using the SMOTE technique, efficacy of the models can be impeoved. To further identify and include only the features that are significant or affect the severity of the accident, feature selection methods such as SelectKBest, Sequential forward selection, and feature importance utilizing model coefficients of the Logistic Regression can be used. Finally, the strategy of applying stacking ensembling technique can be applied to make better forecasts.

In the study, the unbalanced dataset is handled using the SMOTE algorithm, and then features that significantly affect accident severity are found using feature selection algorithms like SelectKBest, Sequential Feature Selector, Recursive Feature Elimination, and feature selection using model coefficients. The study uses machine learning methods including k-Nearest Neighbors, Naive Bayes, and Random Forest classifier to forecast the severity of the accidents. Further the study propose improvement in performance by building a Stacking Ensemble classifier that uses diverse base learners and a new meta learner to create the final predictions.

III. METHODOLOGY

The method followed in the study is comparative analysis of machine learning model on unbalanced and balanced dataset for predicting accident severity. Fig. (1a and 1b) describes the methodological flow of steps for building machine learning models for unbalanced (Fig. 1a) and (Fig. 1b) for balanced datasets.

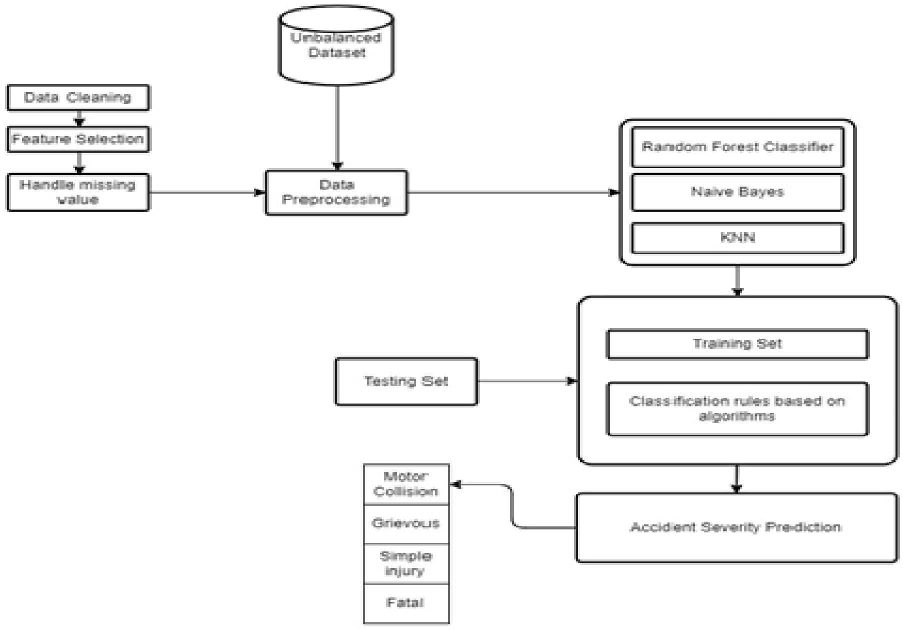

Fig. 1.a. Methodology for Unbalanced dataset

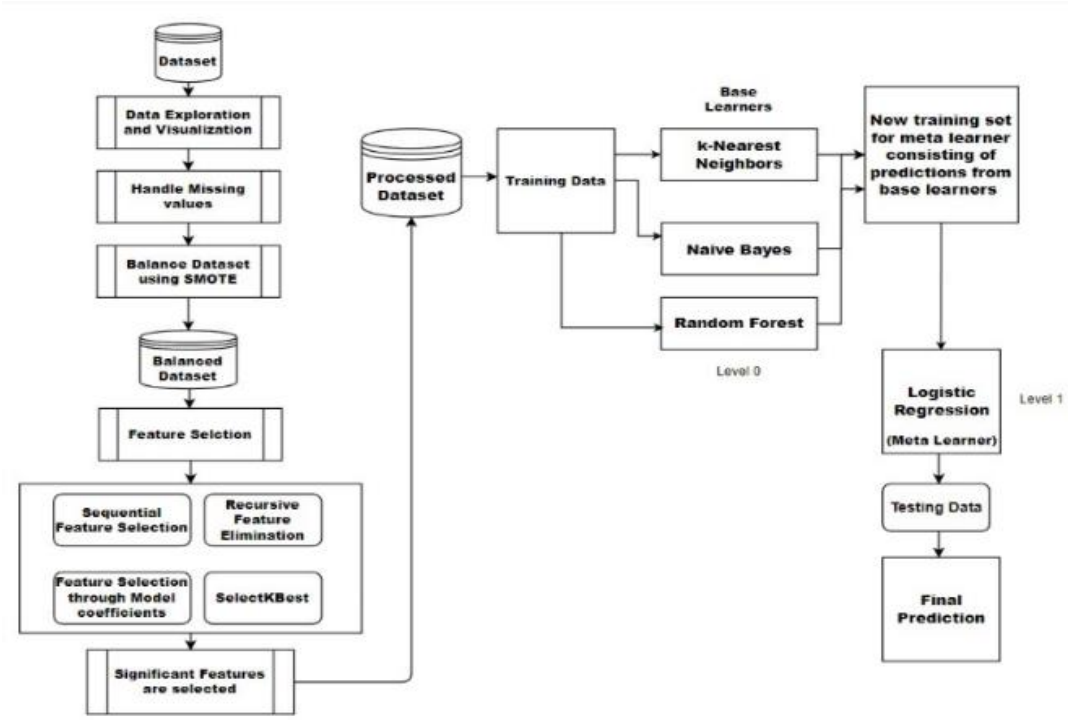

Fig. 1.b. Methodology for balanced datasets

A. llection of Data

The dataset was collected from source mentioned in literature [23]. The dataset contains about 2.8 million records each with 47 features such ID, Severity, Start_Time, End_Time, Start_Lat, Start_Lng, End_Lat, End_Lng, Distance(mi), Description, Number, Street, Side, City, County, State, Zipcode, Country, Timezone, Airport_Code, Weather_Timestamp, Temperature(F), Wind_Chill(F), Humidity(%), Pressure(in), Visibility(mi), Wind_Direction, Wind_Speed(mph), Precipitation(in), Weather_Condition, Amenity, Bump, Crossing, Give_Way, Junction, No_Exit, Railway, Roundabout, Station, Stop, Traffic_Calmin, Traffic_Signal, Turning_Loop, Sunrise_Sunset, Civil_Twilight, Nautical_Twilight, Astronomical_Twilight. The size of the dataset is 1.15GB. There are 13 features of datatype bool, 13 features of data type float, 1 feature with datatype int and 20 features have object data type. The severity feature is defined by 4 labels, which indicate the severity of the accident and its impact on traffic. Class 1=Motor collision, Class 2=Simple injury, Class 3=Grievous, Class 4=Fatal. Table 1 shows the dataset description.

B. Data Exploration and Preprocessing

Through data exploration we could get more insights about the dataset that we are working on. We were able to find the number of features that contain null values and according to the type of feature we were able to handle the null values by filling them with either median, mean values and in case of time series data, the null values were filled using the bfill and ffill methods. By plotting the heatmaps, we were able to understand the correlation between the features and this helped us in the feature selection process as well. For feature selection, we have used filter, wrapper and embedded methods and from those results we were able to select the most important features that will be further used in the model fitting purpose. The phrase "feature importance" refers to a collection of methods for rating input qualities in a predictive model, indicating the relative weight given to each characteristic in the creation of a prediction. The feature importance of a predictive model can be used to enhance it. To achieve this, choose which features to maintain and which to discard based on the importance scores.

C. Feature Selection using SelectKBest algorithm

To extract the top features from a dataset, the Scikit-learn API class called SelectKBest was used. The SelectKBest method selects features based on the feature with the highest k score. Choosing the best features is essential when getting ready a big dataset for training. It helps cut down on training time and the removal of less important data. Here, the study applied the chi2 scoring function. The univariate feature selection approach is the foundation of the SelectKBest algorithm. This selection approach compares a single input variable to the target variable to alculate the statistical measures. Fig. 2. depicts the steps in SelectKBest algorithm. Initially, there were 47 columns in the dataset, but using the SelectKBest technique, we were able to isolate the 10 most crucial features (Table 2). Because our dataset primarily comprises of categorical data, the chi-squared test was the most commonly utilized correlation test. It investigates whether there is a statistically significant difference between the observed and expected frequencies of two category variables.

Fig. 2. Feature selection using SelectK Best algorithm

Table 2 Feature selection using SelectK Best

Features	Chi^2 test Score
Distance(mi)	117707.342
Humidity(%)	101199.794
Wind_Speed(mph)	55621.063
Traffic_Signal	38424.053
Crossing	26485.062
Junction	15888.792
Side	9102.546
Sunrise_Sunset	2791.484
Visibility(mi)	1689.014
Station	1040.8219

Sequential Feature Selection: In general, algorithms for selecting features sequentially are a subset of wrapper methods that successively add and remove features from the dataset. It may occasionally evaluate each feature separately and select M features from N available features based on the individual ratings. To gain the most crucial features for our solution, we used the sequential forward selection. In the forward selection version, characteristics are gradually added to an empty collection of features until the addition of new features does not reduce the criterion. We were able to extract the following features by putting the aforementioned technique into practice: distance (miles), humidity (percent), pressure (in), visibility (miles), wind speed (mph), precipitation (in), crossing, give-way, junction, and station.

Recursive Feature Elimination: The external estimator that gives weights to features, recursive feature elimination (RFE) seeks to identify features by periodically taking into account smaller and lesser sets of features. The estimator is trained on the original set of characteristics to estimate the relevance of each feature, which is subsequently determined by each individual attribute. Then, the least important elements are dropped from the list as it stands. The method is then iteratively applied to the pruned set until the right amount of features have been chosen. RFE is a feature selection algorithm of the wrapper type. The following features were obtained using this feature selection algorithm. Distance(mi), Temperature(F), Pressure(in), Wind Chill(F), Humidity(percent), Wind Speed(mph), and Precipitation(in) are the many weather variables.

Given that we are working with a classification algorithm, it made sense for us to apply a logistic regression model to determine the significance of the features. The coefficients are kept in the coef_ property once the model has been fitted. If the coefficient for a characteristic is a significant negative or positive value, this suggests that the feature had some bearing on the prediction. The prediction is unaffected if the coefficient for a feature is zero, though. In Fig. 3, the feature scores are displayed.

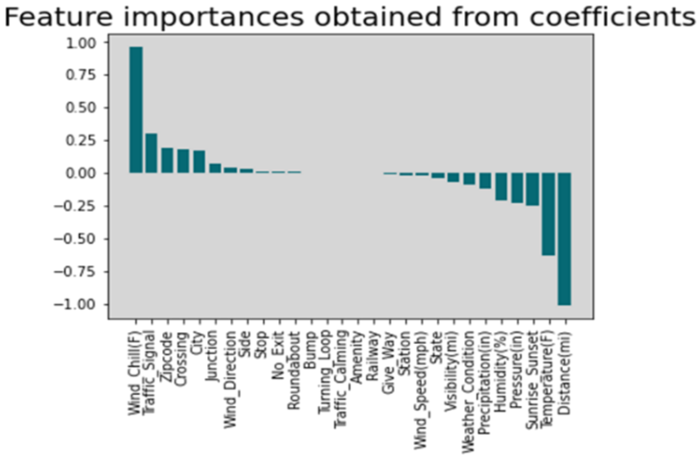

Fig. 3. Feature importance using coefficients of Logistic Regression

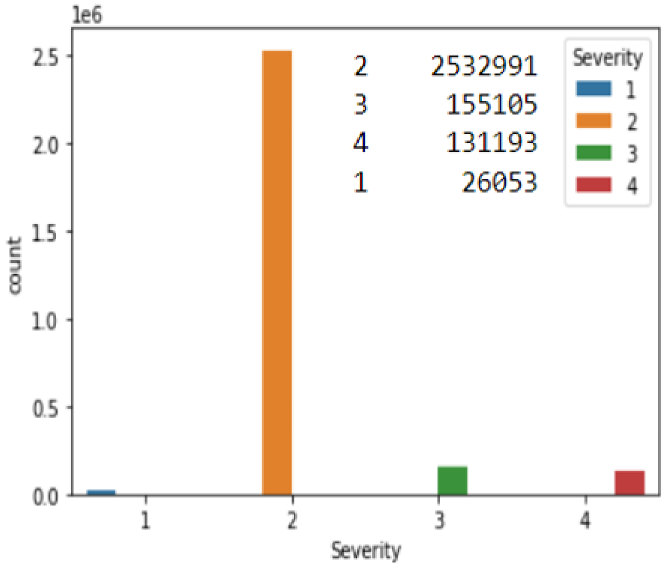

Fig. 4. Distribution of classes

The collection of features on which our model would be trained was created by merging the features obtained from each of the feature selection techniques; these features were the most significant features. The final features chosen were "Severity," "Distance(mi)," "Side," "Temperature(F)," "Wind Chill(F)," "Humidity(%")," "Pressure(in)," "Visibility(mi")," Wind Speed(mph")," Precipitation(in),"Crossing”, "Junction," "Station," "Stop," "Traffic Signal," and "Sunrise" and "Sunset".

D. Handling Imbalance in the Dataset

Fig. 4. shows that the dataset has about 2.85 million records, of which 2.5 million records are of the accident severity type simple injury (severity level=2), which is the majority class, and only 26,053 records are of the minority class Motor collision (severity level=1), which accounts for less than 10% of the dataset's total size. From this, the study infers that there is a slight imbalance in our dataset. To further balance our dataset, two additional procedures were used. From the Fig. 4. it is evident that the dataset contains about 2.85 million records, out of which 2.5 million records belong to the accident severity type simple injury (severity level=2), this is the majority class and the minority class Motor collision (severity level=1) has only 26,053 records which is less than 10% of the total size of the dataset, from this we could conclude that there is a mild imbalance in our dataset.

Further, two approaches were taken to balance the dataset. The first strategy involved using the random oversampling technique, but this method produced redundant data and did not add any new variety to the balanced dataset; instead, the new records were simply duplicates of the previous data points, which did not help to enhance the performance of our classification models. Finally, we used the SMOTE approach [12], which added a wide range of additional data points to the dataset while also balancing it so that each severity category had an equal number of samples. The balanced dataset then contained 3.3 million data records.

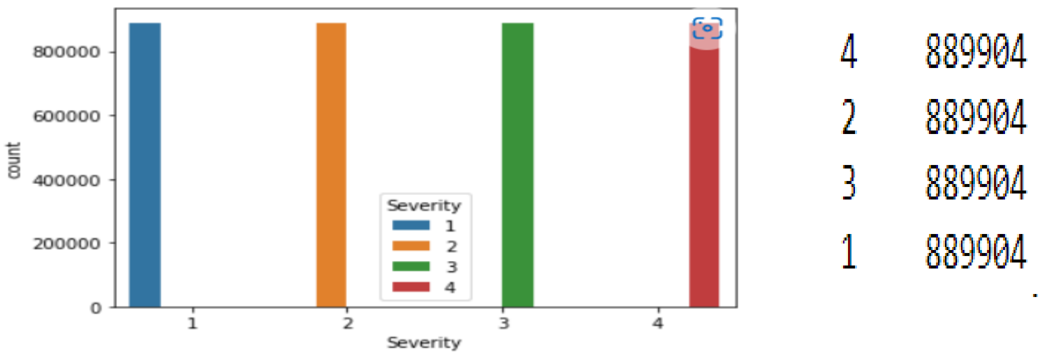

Fig. 5. Distribution of Severity levels in balance dataset

E. Machine learning algorithms for prediction

The study deals with multi classification. A variety of machine learning algorithms were used to predict the severity of the accident. The severity of the accident can be classified into 4 different categories which are in the range of 1 to 4. These values can be further encoded to 1: Motor Collision, 2: Simple injury, 3: Grievous, 4: Fatal.

K-Nearest Neighbors was the initial algorithm that was used with the optimized value of k being set to 5.

Followed by an effective probabilistic machine learning approach for multi-class classification i.e. naive Bayes. The RF was then used in the study, which is a form of supervised learning technique in which a number of decision trees are trained on different subsets of the original dataset and the random forest classifier takes the average of all the decision trees and utilizes it to enhance the accuracy of its prediction. After fitting many decision tree classifiers to distinct dataset subsamples, a random forest meta estimator uses averaging to improve projected accuracy and avoid overfitting with hyperparametric optimization. Given that the study required fine-tuning of a greater number of parameters and required lengthy training periods for each model fitting, RandomizedSearchCV was chosen since it is less computationally expensive and efficient. Following completion of hyperparameter tuning using the

RandomizedSearchCV, we achieved the best parameters, which are listed below:

Optimal Parameters n estimators = 128, min samples split = 6, min samples leaf = 3, max features = auto, max depth = 40, bootstrap = false.

<u>Stack Ensemble technique:</u> Stacking is an ensemble machine learning approach that use a meta learner algorithm to efficiently integrate the predictions of the base learners. It will make predictions that perform better than any of the individual base models employed in the ensemble technique by utilizing the prediction abilities of the individual base learners. In stacking, a single meta learner is used to integrate the predictions from the basis models using heterogeneous meta learners. The base learners' predictions on out-of-sample data are used to train the meta learner. Predictions from the base learners and anticipated results are used as input and output pairs in the training dataset that is used to fit the meta learner. The basis learners utilized here are sophisticated; Naive Bayes, Random Forest, and KNN serve as our base learners, and Logistic Regression serves as the meta learner for the prediction. We were able to surpass the accuracy of our foundation learners using this method. The Stacking Ensemble Classifier's general architecture is depicted in Fig. 6. The stacking ensemble method outperformed our base learners in terms of accuracy and a number of other criteria. Therefore, using the stacking ensemble technique seemed appropriate.

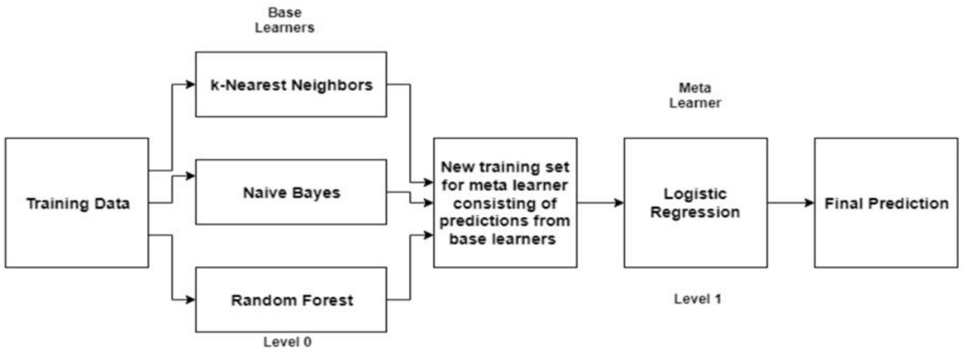

Fig. 6. Architecture of stacking ensemble classifier

IV. RESULTS AND DISCUSSION

This section presents and discusses the experiments and conclusions for three different classifiers and a stacking model: Decision tree - Random Forest, K Nearest Neighbors, Naive Bayes, and stacked ensemble model. Various comparisons and analysis for both balanced and unbalanced datasets were addressed to establish which of the four techniques performed best at forecasting accident severity.

A. Results for unbalanced data.

The comparative results for unbalanced data set are depicted in Table 3. It is evident that although the accuracy is high, the other metrics such as precision, recall and f1 score is very low. Here RFC outperforms and naive bayes underperforms. This makes the model unreliable and since the data is moderately unbalanced it creates unconscious bias.

Table 3 Results for unbalanced data

Methods	Accuracy	Precision	Recall	F1-Score
Naive Bayes	78.5	0.3	0.36	0.28
KNN	88.5	0.52	0.32	0.36
RFC	90.1	0.71	0.44	0.50
Stacking Ensemble	89.1	0.55	0.4	0.45

B. Results after balancing the data using SMOTE

The comparative results for balanced data set are depicted in Table 4. It is clear that balanced data models have a higher accuracy and precision. Furthermore, objective is to improving Accuracy and other performance metrics of the algorithms using stacking ensemble, which is shown in Fig. 8.

Table 4 Result metrics for Balanced Data

Methods	Accuracy	Precision	Recall	F1-Score
Naive Bayes	46.5	0.48	0.47	0.41
KNN	95.3	0.95	0.95	0.95
RFC	95.45	0.95	0.95	0.95
Stacking Ensemble	96.92	0.97	0.97	0.97

C. Comparing the model results of Balanced and Unbalanced data.

The findings of a precise comparison of the employed algorithms showed (Table 5) that RandomForest had a greater accuracy of 90.1 and precision of 0.71 for unbalanced data and accuracy of 95.45, precision of 0.95 for balanced data as shown in Fig.1 and Fig.2 respectively. Following RFC the next best algorithm found was K

Nearest Neighbors with accuracy of 88.5 for unbalanced and 95.3 for balanced data. Naive bayes algorithm underperformed in both datasets with accuracy and precision of 78 and 0.8 for unbalanced and 46, 0.46 for balanced respectively. It was seen for unbalanced data precision, accuracy recall and f1-score of class 2 was significantly higher when compared to classes 1,3 and 4. This can be interpreted as following: The model was able to make the right prediction for class 2 almost all of the time but failed to predict for the other 3 classes (Classes 1,3 and 4).These findings from Table -1 signifies that model trained on balanced data is more efficient than model trained on unbalanced data as in balanced data the accuracy and precision were almost equally distributed among all classes. Hence unbalanced data modeling was ignored for the final model. Hence all the models of balanced data were used as base learners for the Stacking ensemble model to achieve better performance. Naive bayes, KNN, RFC were used as base learners along with logistic regressions as meta learner to create our stacking ensemble model. With this stacking ensemble our model was able to predict and perform better compared to all the 3 algorithms with accuracy ,precision and f1-scores of 96.92,0.97,0.97 respectively as show in the Fig.2.

Table 5 Results of Performance Measurements with different ML techniques using both balanced and unbalanced Data. (B-Balanced Data / U-Unbalanced Data)

Methods	**Accuracy %**		**Precision**		**Recall**		**F1-Score**	
	B	**U**	**B**	**U**	**B**	**U**	**B**	**U**
Naive Bayes	46.5	78.5	0.48	0.3	0.47	0.36	0.41	0.28
KNN	95.3	88.5	0.95	0.52	0.95	0.32	0.95	0.36
RFC	95.45	90.1	0.95	0.71	0.95	0.44	0.95	0.5
Stacking Ensemble	96.92	89.1	0.97	0.55	0.97	0.4	0.97	0.45

From Fig. 7. it is evident that the Stacking ensemble with 0.99 has the best ROC_AUC_SCORE and hence the overall best predicting model. With respect to Fig. 8 curve nearer to the top left of the graph represents better performance. The classification performance is improved by a higher AUC. When the AUC is larger than 0.7, it shows that the model has excellent pattern recognition prediction ability. From Fig. 7. and Fig. 8. it is seen that RFC and Stacking ensemble are the top two best performing models with Stacking ensemble models being the best out of these. AUCs provide statistical evidence of the Stacking ensemble's superior classification ability in this investigation.

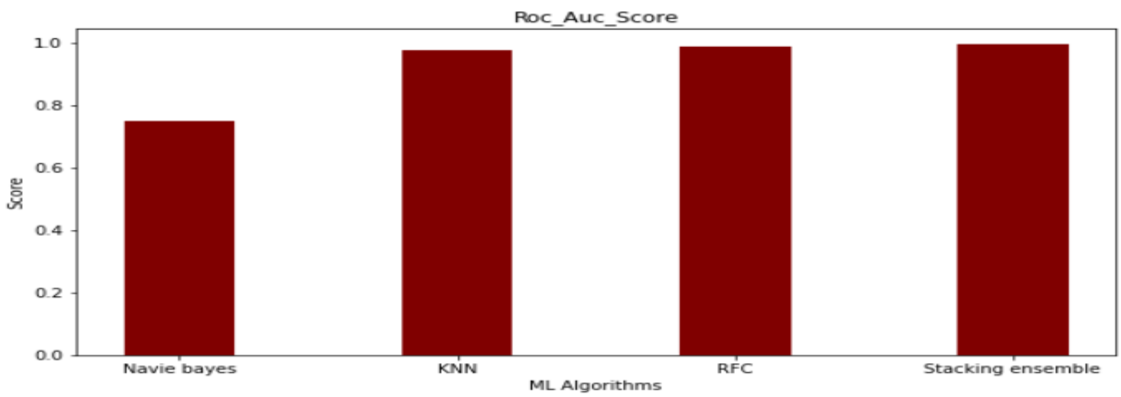

Fig. 7. ROC_AUC_SCORE for different ML Algorithms

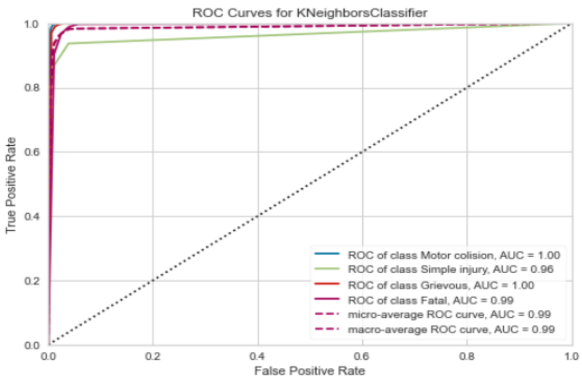

Fig. 8.a. ROC curve for KNN

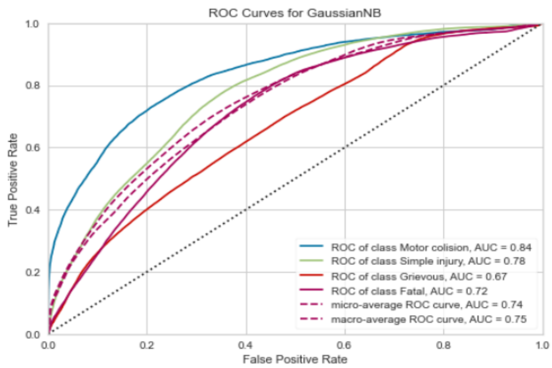

Fig. 8.b. Roc curve for Naive Bayes

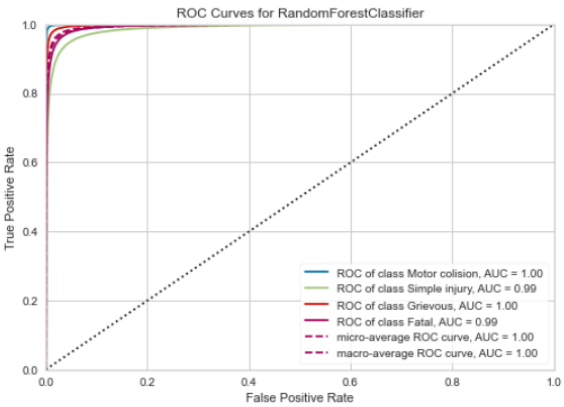

Fig 8.c. ROC curve for RFC

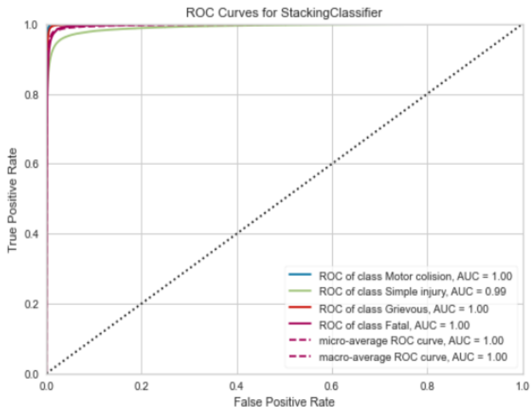

Fig 8.d. ROC curve for Stacking Classifier

Fig. 8. ROC curves of different ML Algorithms.

Figures 8(a), (b), (c), (d) describes the ROC_AUC curve for various classes such as motor collision, simple injury, grievous injury, and fatal. In fig. 8(a), the ROC_AUC achieves a value=1 using KNN classifier. For Naive Bayes classifier in fig. 8(b), the average ROC_AUC value is 0.74, whereas Random-Forest Classifier model achieves 0.99 as average accuracy as shown in fig. 8(c). The Stacking Classifier of fig. 8(d) achieves the maximum value=1 for all the classes.

V. Conclusion

In order to build classifiers that were accurate, precise, and reliable, this study examined how well three machine learning algorithms worked in conjunction with a stacking ensemble model. Includes the Naive Bayesian Classifier, Random Forest, and K Nearest Neighbors in addition to the Stacking Ensemble Model. Although the accuracy of the unbalanced data was equivalent, other performance indicators, including as the AOC, F1, and memory scores, had much lower values and therefore the data set was made balanced using SMOTE, to produce more precise and accurate findings. Performance metrics such as AUC ROC, F1-Score, precision, recall, and accuracy suggested that the stacking ensemble model performed better than the other models. This model can predict accident severity with a 96.92% accuracy, according to the study sample.

The outcomes can be incorporated into an existing real-time accident risk prediction model or used to create a new real-time severity accident risk prediction model to predict an accident's severity in advance, helping to prevent or take precautions while also providing the victims with the necessary and prompt aid. The government and other local authorities can implement the necessary actions with the aid of this model by positioning emergency vehicles such as ambulances, police cars, and fire trucks for prompt assistance. This can be integrated with the current vehicle safety systems. Further research can be done on the intricate relationships between a few key criteria of accident severity.

REFERENCES

[1] "Road traffic injuries," World Health Organization, https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries (accessed Jun. 20, 2022).

[2] R. D. Shirwaikar, U. D. Acharya, K. Makkithaya, M. Surulivelr, and L. E. Lewis, "Machine learning techniques for neonatal apnea prediction," *Journal of Artificial Intelligence*, vol. 9, no. 1–3, pp. 33–38, Dec. 2015. doi:10.3923/jai.2016.33.38

[3] R. D. Shirwaikar *et al.*, "Optimizing Neural Networks for medical data sets: A case study on neonatal apnea prediction," *Artificial Intelligence in Medicine*, vol. 98, pp. 59–76, Jul. 2019. doi:10.1016/j.artmed.2019.07.008

[4] R. D. Shirwaikar, D. Acharya U, K. Makkithaya, S. Mallayaswamy, and L. E. Lewis, "Estimation of caffeine regimens: A machine learning approach for enhanced clinical decision making at a neonatal intensive care unit (NICU)," *Critical Reviews in Biomedical Engineering*, vol. 46, no. 2, pp. 93–105, 2018. doi:10.1615/critrevbiomedeng.2018025933

[5] Sahare, M and Gupta, H, "A review of multi-class classification for imbalanced data," *International Journal of Advanced Computer Research 2.3* , 2012, p. 160.

[6] A. Parmar, R. Katariya, and V. Patel, "A review on Random Forest: An ensemble classifier", *International Conference on Intelligent Data Communication Technologies and Internet of Things (ICICI) 2018*, Dec. 2018. doi:10.1007/978-3-030-03146-6_86

[7] Reid and Sam, *A Review of Heterogeneous Ensemble Methods*. University of Colorado , Boulder : Department of Computer Science, 2007.

[8] L. G. Cuenca, E. Puertas, N. Aliane, and J. F. Andres, "Traffic Accidents Classification and Injury Severity Prediction", 3rd IEEE International Conference on Intelligent Transportation Engineering (ICITE) 2018, pp. 52–57 doi: 10.1109/ICITE.2018.8492545

[9] Zhang. J, Li. Z, Pu. Z, and Xu. C, "Comparing Prediction Performance for Crash Injury Severity Among Various Machine Learning and Statistical Methods," *IEEE Access*, vol. 6, pp. 60079–60087, 2018. doi:10.1109/ACCESS.(2018).2874979

[10] M. F. Labib, A. S. Rifat, M. M. Hossain, A. K. Das, and F. Nawrine, "Road Accident Analysis and Prediction of Accident Severity by Using Machine Learning in Bangladesh", 7th International Conference on Smart Computing & Communications (ICSCC) 2019, pp. 1–5. doi: 10.1109/ICSCC.2019.8843640.

[11] M. Sameen and B. Pradhan, "Severity prediction of traffic accidents with recurrent neural networks," *Applied Sciences*, vol. 7, p. 476, 2017. doi:10.3390/app7060476

[12] Li. Y. Sameen, Ma. D. Pradhan, Wang. Y, Zeng. Z, and Zhu. M, "Identification of significant factors in fatal-injury highway crashes using genetic algorithm and neural network," *Accident Analysis & Prevention*, no. 111, pp. 354–363, Feb. 2018. doi:0.1016/j.aap.2017.11.028

[13] Abellán. J, López. G, and Oña. A, "Analysis of traffic accident severity using Decision Rules via Decision Trees," *Expert Systems with Applications*, vol. 40, no. 15, pp. 6047–6054, 2013. ISSN 0957-4174. doi: https://doi.org/10.1016/j.eswa.2013.05.027.

[14] M. Yan and Y. Shen, "Traffic accident severity prediction based on random forest," Sustainability, vol. 14, p. 1729, Feb. 2022. doi:10.3390/su14031729

[15] R. E. Al Mamlook, K. M. Kwayu, M. R. Alkasisbeh, and A. A. Frefer, "Comparison of Machine Learning Algorithms for Predicting Traffic Accident Severity", IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT) 2019, pp. 272–276. doi: 10.1109/JEEIT.2019.8717393.

[16] C. Dong, C. Shao, J. Li, and Z. Xiong, "An improved deep learning model for traffic crash prediction," *Journal of Advanced Transportation*, vol. 2018, pp. 1–13, 2018. doi:10.1155/2018/3869106

[17] 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, pp. 321–357, 2002.

[18] H. Ren, Y. Song, J. Wang, Y. Hu, and J. Lei, "A Deep Learning Approach to the Citywide Traffic Accident Risk Prediction", 21st International Conference on Intelligent Transportation Systems (ITSC) 2018, pp. 3346–3351. doi: 10.1109/ITSC.2018.8569437

[19] D. Santos, J. Saias, P. Quaresma, and V. B. Nogueira, "Machine learning approaches to traffic accident analysis and Hotspot Prediction," *Computers*, vol. 10, p. 157, 2021. doi:10.3390/computers10120157

[20] A. Theofilatos, D. Graham, and G. Yannis, "Factors affecting accident severity inside and outside urban areas in Greece," *Traffic Injury Prevention*, vol. 13, no. 5, pp. 458–467, 2012. doi:10.1080/15389588.2012.661110

[21] Chang. L and Chen. W, "Data mining of tree-based models to analyze freeway accident frequency," *Journal of Safety Research*, vol. 36, no. 4, pp. 365–375, 2005. ISSN 0022-4375 doi:https://doi.org/10.1016/j.jsr.2005.06.013

[22] Hyunjin. K, Park. J, and Lee. Y, "Stacking ensemble technique for classifying breast cancer," Healthcare Informatics Research , vol. 25, no. 4, pp. 283–288, 2019.

[23] M. Sobhan, M. H. Samavatian, S. Parthasarathy, and R. Ramnath, "A Countrywide Traffic Accident Dataset," arXiv preprint arXiv:1906.05409, 2019