INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
Online ISSN 1841-9844, ISSN-L 1841-9836, Volume: 18, Issue: 5, Month: October, Year: 2023
Article Number: 5317, https://doi.org/10.15837/ijccc.2023.5.5317

# Intelligent Model for Avoiding Road Accidents Using Artificial Neural Network

Manoj Kushwaha, M.S. Abirami

**Manoj Kushwaha**
Department of Computational Intelligence
SRM Institute of Science and Technology, KTR Campus, Chengalpet, Tamilnadu, 603203, India
mk3342@srmist.edu.in

**M.S. Abirami**
Department of Computational Intelligence, SRM Institute of Science and Technology, KTR Campus, Chengalpet, Tamilnadu, 603203, India
Corresponding author: abiramim@srmist.edu.in

**Abstract**

Accidents typically occurred on roads, resulting in significant societal losses. Road accidents are a worldwide issue that result in the loss of precious human lives and property. The purpose of this paper is to create an intelligent system-based on Machine Learning model for avoiding road accidents, as well as a system that effectively reduces road accidents severities. The Artificial Neural Network (ANN) algorithm, along with others such as Logistic Regression (LR), Decision Tree (DT), K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Nave Bayes (NB), Stochastic Gradient Descent (SGD), Random Forest (RF), Gradient Boosting (GB), and AdaBoost, is used to create an intelligence system. Many driving collaborator procedures, installed in a few vehicles, assist drivers in avoiding vehicle crashes by providing early cautioning messages. The intelligence road crash avoidance system model is built on dataset of 29 columns and 1048575 rows. Pre-processing, feature selection, and feature extraction performed with the help of heat map and correlation matrix are used to select features. Linear Discriminant Analysis (LDA) is used for feature extraction. The testing dataset revealed that the proposed ANN method outperforms other algorithms with an accuracy of 0.856. Intelligent systems aid in the prevention of traffic accidents, which aids police officers and researchers in developing new policies.
**Keywords:** Machine Learning, Internet of Things, Intelligent System, Artificial Neural Network, Linear Discriminant Analysis.

## 1 Introduction

Road is a common location for both pedestrians and drivers, and it is where road accidents occur. Road accidents resulted in severe injuries and the loss of human lives, animals, and property. Road accidents kill 1.3 million people per year, with an average of 3,287 deaths per day and between 20 and 50 million wounded or incapacitated [28]. World Health Organization, report estimated yearly majority people die in road accidents, between the ages of 19 and 29 [5]. The cost of road traffic fatalities and property damage is disproportionately high in developing countries [11]. A road accident occurs when one vehicle collides with another vehicle, a pedestrian, or a stationary object such as a tree, pole, building, or divider, resulting in significant injuries or death [20]. Due to rapid growth of population, demand of vehicles is also increasing. Researchers focused on the Machine Learning (ML) algorithm-based model which help to predict the severity of accidents. Severe injuries and fatal person from road accidents could be rescue by rapid availability of medical units. Road accidents puts burden on hospitals and affect the health care system [11]. Overspeeding, reckless driving, weather conditions, road surface, and other factors contribute to severe injuries in road accidents [20], [30].

Various well-known existing methods for prediction analyzed for avoiding and preventing the road accidents. Truck crashes typically result in greater economic losses and crash severity when compared to other types of vehicle crashes [8]. The eighth most leading cause of death for people of all ages is traffic accidents [9].

Recently transportation agencies around the world working on to discover accident risk and prediction approaches along with establish standards and working events for road investigations [19]. Transportation experts and practitioners have worked extensively to enhance road safety. Work on the cause, various road accident factors, and accident zone area, and create a hybrid ML model. AI can also access drivers, vehicles, and road data using embedded sensors. Develop an intelligent road avoidance system with the help of AI and ML algorithms to save precious human lives from unpredicted deaths from road accidents. The goal is to create an AI-based ML model for a road accident avoidance system that effectively reduces road accidents using ML and (Internet of things) IoT techniques. To avoid road accidents, we used various ML algorithms, trained our model, and applied AI-based systems to it. Many approaches, such as neural networks, support vector machines, and decision trees, are employed to forecast accidents [14]. This model intends to create an intelligent accident avoidance system that can forecast road accidents for various stages of driving mechanization, such as manual driving, basic driver assistance, and partial driver assistance. Most ML techniques are used to test the proposed intelligent accident avoidance system, including KNN, Decision Tree (DT), Support Vector Machine (SVM), Stochastic Gradient Descent (SGD), Random Forest (RF), Artificial Neural Network (ANN), Naive Bayes (NB), Logistic Regression (LR), Gradient Boosting (GB), and AdaBoost.

The following are the primary contributions of the proposed model:

- The model is cost effective and was trained using a hybrid of ANN and other algorithms. The system is applicable to both human-driven and self-driving vehicles.
- A novel ensemble of ML Models for predicting accidental severity has been proposed.
- A comparison of different Ensemble learning classifiers like LR, KNN, SVM, NB, DT, RF, SGD, ANN, GB, AdaBoost.
- The effect of crucial variables on estimating metrics such as accuracy, precision, and recall, which were used as input parameters for evaluating the severity of road accidents, was investigated.
- Because reducing the number of input variables enhances efficiency, it also reduces data gathering costs.
- Training on merged datasets with fewer missing values and outliers can improve prediction accuracy even further.

Database is pre-processed to drop absent data, normalization, and standardisation before being fed into ML algorithms [37, 38]. Following that, the dataset is divided into two sets: training and testing. ANN and other models are trained using the training set. Intelligent systems may be used to anticipate road collisions at each degree of driving mechanisation, from underdeveloped driver assistance at level 1 to complete mechanisation at level 5. Models predicted automotive collision levels with an accuracy of 85.6% using the testing dataset. The remaining section discussed related works, methodology, result and conclusion.

## 2 Related Work

Several road assistant systems have been proposed in the literature to minimize car accidents.

Guo Xie et al. [1] proposed a model that would reduce the probability of traffic accidents, extract the key influences on them, and provide the foundation for unmanned vehicle decision-making.

Yuchuan Fu et al. [2] discussed various collision avoidance methods. The collision avoidance (CA) phase is divided into three phases: prediction, risk appraisal, and reaction. These processes are used to interpret sensor data, recognise possible hazards, and respond in time to avoid or minimise them. In general, CA algorithms can be executed on a vehicle, an edge server, or a central cloud server.

Yohan Chang et al. [3] introduced a two-part model that 1) uses LR models to assess CNC risk based on pre-incident characteristics, and 2) uses ML approaches to classify and forecast CNC incidents.

Natalia Selini Hadjidimitriou et al. [4] suggested a non-linear ML system may be used to build a useful and reliable intelligent system for the severity categorization of accidents involving two-wheeled vehicles based on operational features that can be reliably gathered by sensors or other on-board equipment.

Alexandre Moreira Nascimento et al [6], suggested a preliminary insight regarding the significance of having a major safety agenda for future AI-based AV system research.

Vivek Singhal Et al. [10] focuses on developing a road vehicle-train accident avoidance system for unmanned railway level crossings using artificial intelligence-enabled risk prediction analysis. Using artificial intelligence-enabled risk prediction analysis.

Jaehwan Kim et al. [12] proposed a collision risk assessment method that quantitatively assesses collision risks for a collection of local path choices using lane-based probabilistic motion prediction of nearby cars.

Yannis George et al. [13] investigated three forms of accident severity: The total number of cars involved divided by deaths; the total number of serious injuries divided by the total number of vehicles involved; and the total number of mild injuries divided by the total number of vehicles involved.

Daichi Nozaki et al. [15] proposed a method for informing and notifying automated driving trucks and crane cars of information that enables truck and crane drivers to prevent collisions. Because multiple cars equipped with 4K cameras are simultaneously streaming pictures via the 5G uplink, uplink traffic may get overloaded.

Surraiya Islam Tonni et al [16] proposed an AI-based driver vigilance system to help drivers avoid accidents. The system uses the Convolutional Neural Network algorithm to detect driver drowsiness from the dash camera; the two-layer long short-term memory algorithm to detect anomalies in the heartbeat; and GPS and the front camera to detect over-speed. The proposed model integrates drowsiness, heartbeat, and speed variables and generates alerts and controls brakes as needed.

Luisito L. Lacatan et al. [17] employ a deep learning, high accuracy brake light recognition system to focus the YOLOv3 algorithm on a viable car safety feature and rear-end accident avoidance. It can detect the car ahead of it at any time of day or night as long as it can see the brake lights of the vehicle ahead of it, even if they are not in the same lane.

Abirami M.S. et al. [27] investigated how socioeconomic and sociocultural factors, such as the size of one's friend group, the number of friends who drink and smoke, information about one's parents, and so on, play a role in the initiation and cultivation of addictive behaviours and used a machine learning approach to predict the early onset of such behaviors. Using our multi-classifier prediction technique, we trained two classifiers to predict if a person drinks or smokes. reduced computational complexity from big label subsets at the same time and met the desire for a focus with multi-label training data. Different summary of road accidents avoidance related research shown in Table 1.

# 3 Methodology

System used different ML algorithms and tried to reduce the road accidents, severe injury and death cases. Following methods were used to develop intelligent model.

## 3.1 Data Collection

I make use of information on car crashes gathered by the United Kingdom (UK) Department of Transport between 2005 and 2015 [35, 36]. There are three separate CSV files included in this dataset, accidents, causalities and vehicles. The main table is Accidents and has references by Accident Index to the casualties and vehicles tables.Using a CSV library python's csv module to read the original csv files and filter the rows based on our desired attributes (no missing value) by using filtered data.append (row). Created a new combined dataset from the existing three datasets. The features include a variety of information, such as Accident Severity (further classified as fatal, serious, slight), Number of Vehicles which associated with road accidents, Number of Casualties, Date, Day of Week which covers around 94 districts or areas in UK, Speed limit, Junction Detail, Junction Control, Light Conditions, Weather Conditions, Road Surface Conditions, Urban or Rural Area, Vehicle Reference, Casualty Reference, Casualty Class, Sex of Casualty, Age of Casualty, Casualty Severity, Pedestrian Location, Pedestrian Movement, Casualty Type, Casualty Home Area Type, Vehicle Type, Vehicle Manoeuvre, Sex of Driver, Age of Driver, Engine Capacity (CC), Age of Vehicle. Each feature's description and definition can be found at [35]. There are 1048575 rows and 29 columns in the combined database, which is shown in Figure 1. Also showed all the attributes of dataset with datatypes which is taken from other accident, causality, and Vehicle dataset. Pattern of each attributes, and distribution of dataset., shown in Figure 2.

## 3.2 Data Preprocessing

To begin data pre-processing, the dataset is checked for missing and null values. The typecasting method is used to convert string values into the required numerical format. Date as an example. Then duplicated and outlier values are removed. Normalized the data and Label encoding used for the categorical data to convert into numerical data. Each category assigned a numerical value [20]. Data pre-processing is an important stage in preparing required data. Convert the characteristics from categorical or nominal to distinct values. Some

| Sources | Author | Methods | Observation |
|---|---|---|---|
| [19] | Da-Jie Lin et al. | NB, DT, C4.5, Multilayer perceptron (MLP), Deep Neural Networks (DNN), Deep Belief Network (DBN), Convolution Neural Network (CNN) | Identifying the major environmental factors that influence the occurrence of accidents at crossings Models that predict the likely locations of high-risk accidents Accuracy: 72.64%, CNN and C4.5 |
| [20] | Mahzabeen Emu et al. | SVM, RF, KNN, CNN | Determine whether or not a collision will be fatal. Accuracy: 75%, CNN |
| [21] | Kamalakar Vijay Thakare et al. | K-body proposal algorithm, YOLO-v3 | Perform high-level post-processing to explain the severity and environment of an accident based on automated recognition, location, and description of the accident's occurrence. |
| [11] | Mubariz Manzoor et al. | RF, CNN, Data mining techniques | Presents an ensemble of ML and DL models by merging RF and CNN called RFCNN for the forecast of road accident severity. Accuracy: 0.991 |
| [22] | Choo-Yee Ting et al. | RF, XGBoost, Cart, Neural Network, NB, SVM | Estimation of fatal road accidents found 26 critical features. Accuracy of RF algorithms 95.46%. |
| [3] | Yohan Chang et al. | RF, DNN, Multilayer Feed forward Neural Network, t-SNE | Evaluate risk of CNC and classify and predict CNC (crashes & near crashes) events using ML methodsBest Algorithm: RF, Accuracy: 91.2% for two classes: crash and near -crash |
| [7] | Xiaoyan Shen et al. | Random forest (RF), DNN, (Multilayer Feed forword eural NNetwork, (t-SNE), Xtreme grgadient Boosting (XGBoost) | Simplified areas develop targeted prevention and response strategies to efficiently reduced severity of accidents. Best Algorithm: XGBoost |
| [4] | Natalia Selini Hadjidimitriou et al. | Machine Learning | Recognized the important features that allow to differentiate accident severity. Best Algorithm: SVM, Accuracy: 92.5% |
| [1] | Guo Xie et al. | ANN and PCA | Extracted the significant impact features of road accidents and decrease the likelihood of road accidents. |
| [23] | Sung-Chiang Lin et al. | Hybrid label-based meta-learning with generalised linear mixed model (GLMM), generalised linear mixed model (GLMM) (HybridLBGLM), | Lowered computational complexity from big label subsets while also meeting the requirement for a focus on multi-label training data. |

Table 1: Summary of related research

```
Accident_Severity                    int64
Number_of_Vehicles                   int64
Number_of_Casualties                 int64
Date                                object
Day_of_Week                          int64
Speed_limit                          int64
Junction_Detail                      int64
Junction_Control                     int64
Light_Conditions                     int64
Weather_Conditions                   int64
Road_Surface_Conditions              int64
Urban_or_Rural_Area                  int64
Vehicle_Reference                    int64
Casualty_Reference                   int64
Casualty_Class                       int64
Sex_of_Casualty                      int64
Age_of_Casualty                      int64
Casualty_Severity                    int64
Pedestrian_Location                  int64
Pedestrian_Movement                  int64
Casualty_Type                        int64
Casualty_Home_Area_Type              int64
Vehicle_Type                         int64
Vehicle_Manoeuvre                    int64
Sex_of_Driver                        int64
Age_of_Driver                        int64
Engine_Capacity_(CC)                 int64
Age_of_Vehicle                       int64
dtype: object
```

Figure 1: Road accident Dataset

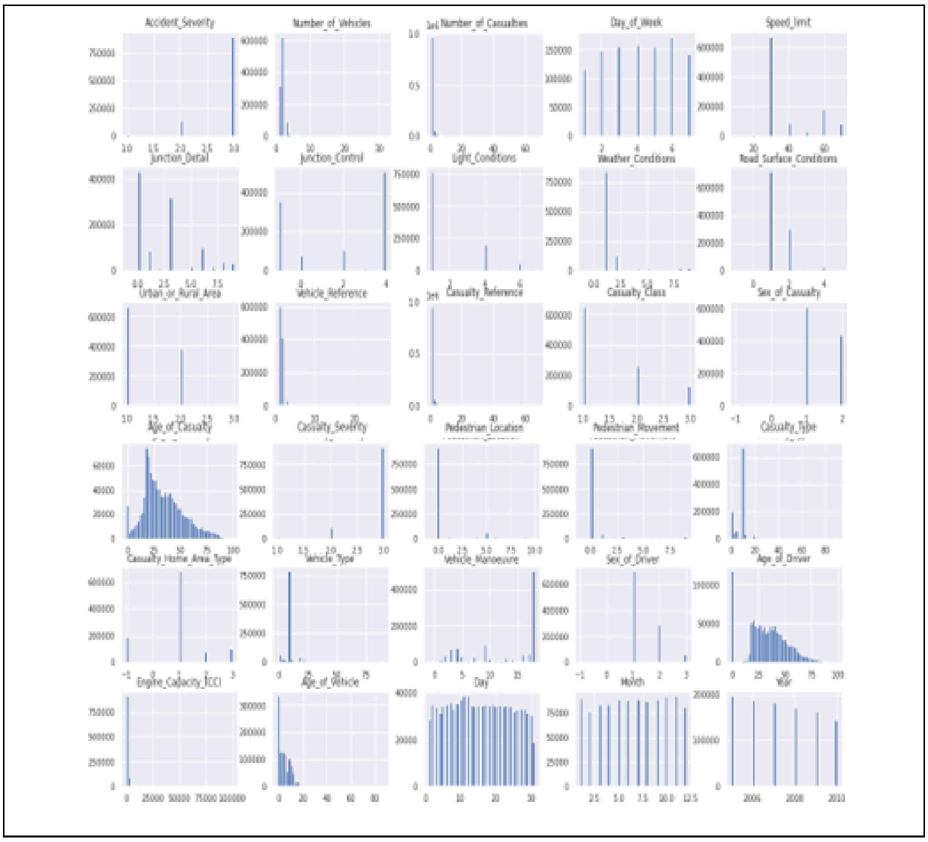


Figure 2: Attributes representation of dataset

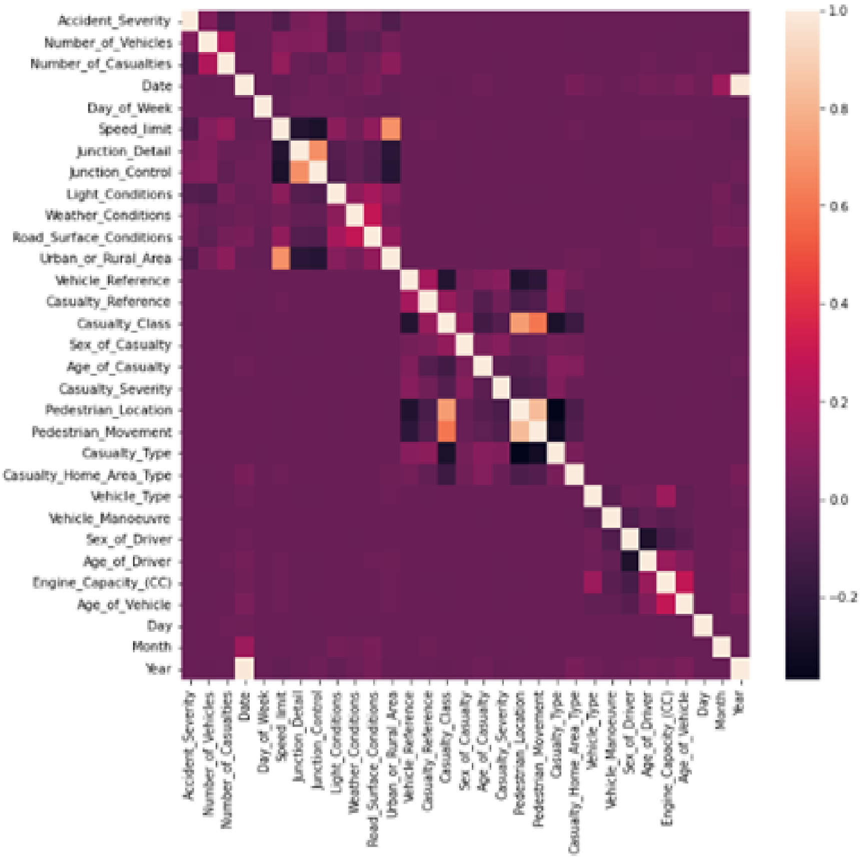


Figure 3: Correlation of different features

data is missing, as is characteristic of all created data. These values are replaced by the column mean. For ML estimators, dataset standardisation is performed, causing ML to disregard the influence of insignificant range characteristics. Further data split into two sections: 70

## 3.3 Feature Selection

Following pre-processing, we must extract required attributes from a set of 29 attributes. The heat-maps correlation matrix shown in Figure 3 was used for feature selection. Only the most relevant variables from the original dataset are retained by Feature Selection. Following the Heat-map, we look for correlations between attributes; if no correlations are found, we discard those attributes and only consider the highly correlated ones. Important data include information that is strongly associated with the target as well as data that is tightly linked to one another [6]. Increasing the value of features directly proportional to increasing the value of variable of target according to this correlation is positive or negative with respect to increasing and decreasing feature values. Data cleaning and feature selection are critical steps in model design. Figure 3 Interested features can be selected automatically or manually using feature selection methods that have a greater influence on road accident attributes. Heatmaps is the process of selecting highly correlated and less correlated attributes of dataset. It defines the relationship in between the attributes of dataset. Here threshold point 1 is the highly correlated attributes which is white colour and black colour is the very less correlated attributes. Highly correlated attributes identified as 'Accident Severity', 'Number of Vehicles', 'Number of Casualties','face Conditions','Urban or Rural Area', 'Vehicle Reference'. Lower correlated attributes as "Age of Driver", "Sex of Driver","Junction Detail","Junction Control", "Pedestrian Location", "Pedestrian Movement", "Engine Capacity(CC)", "Vehicle Reference", "Casualty Home Area Type, "Casualty Reference".

## 3.4 Feature Extraction Techniques

Some commonly used feature extraction approaches are:

### 3.4.1 PCA:

Principal component Analysis (PCA) is an example of an unsupervised ML method used for dimensionality reduction. Dimensionality reduction is the most important work of PCA because it compares several attributes of a dataset that are associated with one another.

### 3.4.2 LDA:

One type of supervised ML that is used to separate two groups or classes. The primary goal of Linear Discriminant Analysis (LDA) is to maximise the distinguishability of the two groups so that we can determine the best way to group them. LDA is similar to PCA in that it reduces dimensionality, but it focuses on improving distinguishability among realised classifications by creating another direct hub and projecting the information focused on that axis.

ML algorithms are at the heart of AI. In general, ML requires two piles of datasets: training and testing. ML algorithms use training datasets to find the best hypothesis and testing datasets to assess the hypothesis's accuracy.

## 3.5 Algorithms

### 3.5.1 Logistic Regression (LR):

The logistic function, which serves as the approach's heart, is named after LR. Statistics devised the logistic function, often known as the sigmoid function, to characterise the features of fast population expansion in ecology that exceeds the carrying capacity of the ecosystem. In eq. (1), it is an S-shaped curve that maps every real-valued integer to a value between 0 and 1, but never exactly between those bounds.

$$\frac{1}{(1+\hat{e}^{-value})} \tag{1}$$

### 3.5.2 K-Nearest Neighbor (KNN):

Solution for classification and regression challenges, best algorithm is KNN and assumes that similar things are close together [29]. It initially determines the value of k as the number of neighbors tracked by the Euclidean distance between them, which is represented in eq. (2) as:

$$d(y_i,\ \hat{y}_i) = \sqrt{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2}$$

In general, k=3 to k=10 is suggested for the trials. Due to its few parameters and straightforward implementation, this approach has been employed in various accident severity-related research investigations to compare with other algorithms [20], [29].

### 3.5.3 Support Vector Machine (SVM):

SVM is a ML technique that may be used for classification or regression problems. It is frequently used in classification problems. Each data item is defined as a point in n-dimensional space in the SVM approach, with the value of each feature being the value of a single coordinate and simply the coordinates of a given observation [20], [22]. The SVM classifier has the advantage of discriminating between the two classes (hyperplane and line) the best. It is a supervised ML tool that is based on statistical theory and the structural risk minimization principle, with the purpose of training nonlinear ML and applying the resultant optimization theory in high-dimensional feature spaces [24], [26], [32]. This machine's basic strategy is to build a large number of separation hyperplanes and then choose the best one to categorise distinct sets.

### 3.5.4 Naive Bayes (NB):

The NB classification approach is a Bayesian probabilistic ML classification technique. In eq. (3), the Bayes' theorem for the classification problem is:

$$P(M|N) = \frac{P(N|M)\,P(M)}{P(N)} \tag{3}$$

where M is the class and N is data. The NB classifier aggregates a Bayesian probability model with a decision algorithm in order to increase the posterior probability [19], [11]. As a result, the goal of the NB classifier is to discover the class (M=m) with the highest probability:

$$\widehat{m_i} = \arg\max_{m} P(m) \prod_{i=1}^{n} P(n_i \mid m) \tag{4}$$

where n is the predictor and m represent class in eq. (4).

#### 3.5.5 Decision Tree (DT):

The DT algorithm is a member of the supervised learning algorithm family. Unlike other supervised learning algorithms, the DT algorithm can also be used to solve regression and classification problems [19], [32].

#### 3.5.6 Stochastic Gradient Descent (SGD):

SGD is a well-known and commonly used ML technique that serves as the foundation for neural networks. A stochastic system or process is one that has a chance of happening. As a result, rather than using the complete data set for each iteration, SGD randomly chooses a few samples. The term "batch" in gradient descent refers to the total number of samples from a dataset used to calculate the gradient for each iteration. In standard gradient descent optimization, such as batch gradient descent, the batch represents the complete dataset. While utilizing the complete dataset is highly valuable for discovering minima in a less noisy and random manner, there is an issue as our dataset becomes enormous.

#### 3.5.7 ANN:

ANN is a message processing model inspired by the neurons systems of humans, specifically the brain. Neurons are the brain's basic building blocks. Each neuron takes input from several other neurons that are linked to it, analyses it, and produces one output that is passed to the next neuron (or neurons).

The columns of the ANN are separated into three layers: the input, the hidden, and the output layer. The ANN receives the training or testing dataset from the input layer. An ANN's hidden layer can have up to L layers, with the output layer providing the model hypothesis [19]. Depending on the problem under consideration, ANN may predict single- or multiclass output. The number of ANN rows equals the number of features plus the bias row in the dataset.

In the hidden layers, an activation function, a type of processing component, models each Artificial Neuron (AN). It is challenging to understand the workings of algorithms like ANNs and SVM and to obtain differences in the impact of various features on accident severity directly Table 2.

#### 3.5.8 Random Forest (RF):

The RF method is an ensemble ML technique that is basically built on a DT [3]. It uses the bagging method to produce numerous subsets of data by randomly picking observations with replacement from the original data sets and then utilising these subsets to build smaller trees. Then, all of these trees were run in parallel, and the final severity prediction of a road accident was calculated by averaging the predictions from all of the DTs. The number of trees, the maximum number of characteristics evaluated to divide a node, and the minimum number of leaves necessary to separate an internal node are three key criteria for enhancing RF prediction accuracy [20], [22], [29], [31], [34].

#### 3.5.9 Adaptive Boosting (AdaBoost):

This is the widely used fundamental ensemble ML method based on boosting [31], [33]. Both the DT and the RF employ a basic learning model. Rather than parallel trees, it develops sequentially. A model is constructed on a portion of the original dataset and then applied to the entire dataset, with equal weights assigned to all data points. The error is computed by subtracting the expected and actual numbers. In a future model, the erroneously predicted data points will be assigned a larger weight. The method is continued until either the error function remains constant or the number of estimators approaches its limit.

| Parameters | Values |
|---|---|
| Number of input layers | 16 |
| Hidden layer sizes | (10, 3) |
| Number of output layer | 3 |
| Learning rate | Constant |
| Max number of iteration (epoch) | 100 |
| Optimizer | Adam |
| Activation Function | Softmax |
| Batch size | 32 |

Table 2: Setup of ANN parameter

#### 3.5.10 Gradient Boosting (GB):

GB is a prominent boosting method. In gradient boosting, each prediction corrects the inaccuracy of its previous. Unlike AdaBoost, the weights of the training instances are not modified; instead, each predictor is trained using the predecessor's residual errors as labels [31]. CART is the base learner for Gradient Boosted Trees (Classification and Regression Trees).

### 3.6 Proposed Model

Dataset is collected from various associated other data, from accident, causality and vehicles data. All data combined in one dataset through python code and made combined dataset. From combined dataset performs data wrangling operation as removing missing value, normalize, and encoding the dataset.

Feature selection perform by Heatmap method, selected highly correlated attributes and drop less correlated attributes from data. With the help of Feature extraction method extract important feature from dataset, our dataset is Supervised dataset so we used LDA method to extract important features and reduce the dimensionality of Dataset.

Further Dataset divided into Training and Testing dataset, applied 10 ML algorithms (LR, KNN, SVM, NB, DT, RF, ANN, SGD, GB, AdaBoost) on testing dataset to compared with performance metrices. It gave the prediction result of severity of road accidents. Accident Severity classified as fatal, serious, slight.

When the car is moving the same location, it will compare with similar features (such as Speed_limit, Age, Location) and it will be experienced with trained algorithms result to avoid road accidents. Schematic proposed model of road accidents avoidance system shown in Figure 4.

There were three severity levels of road accidents: fatal, serious, and slight, as shown in Figure 5, with lesser fatal cases and a higher number of slight severities. The count of slight severity is higher than 800,000, while the count of serious severity is less than 200000.

Accident severity is divided as fatal, serious, and slight according to the 1, 2, or 3 label. Weather conditions are labelled -1, 1, 2, 3,4,5,6,7,8,9, as shown in Figure 6, a graph of the relationship between weather conditions and accident severity.

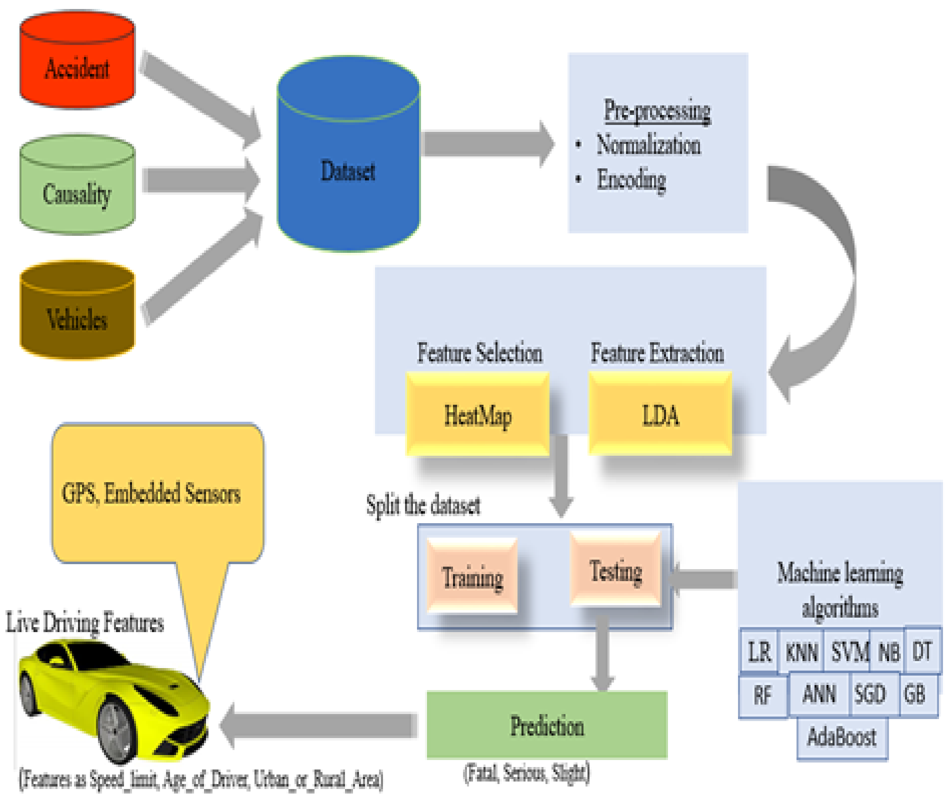


Figure 4: Road Accidents Avoidance System

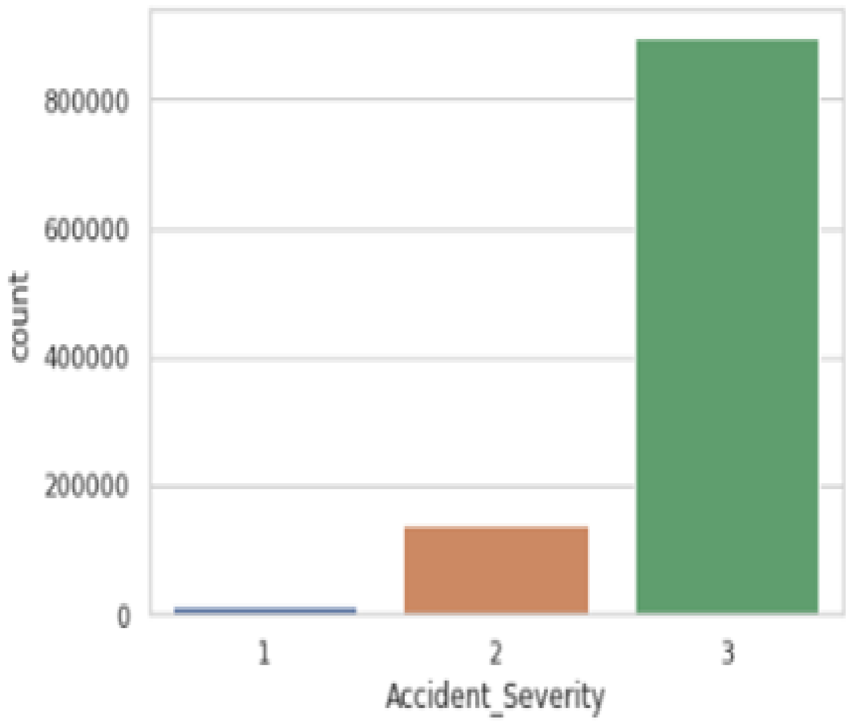


Figure 5: Accident severity levels Fatal, Serious, Slight.

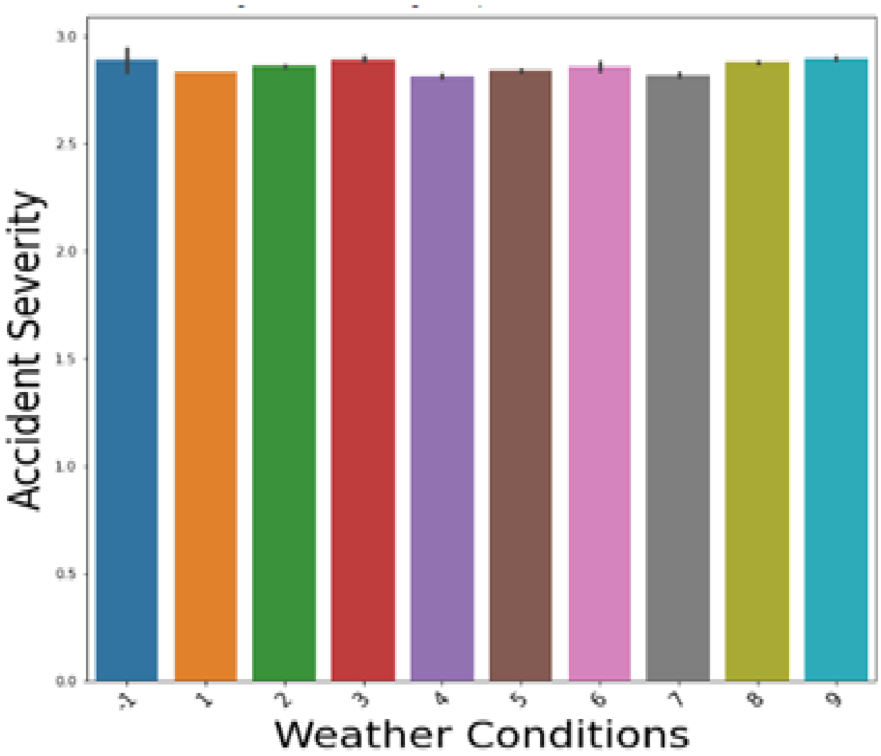


Figure 6: Accident severity vs Weather Conditions.

### 3.7 Algorithm

**Algorithm 1: LDA and Artificial Neural Network**
**Input:** Features, $D_{train,}$
**Output:** $D_{Pred}, D_{Classification}$
I. Performing Feature Scaling
$D_{train,}\ D_{test} \leftarrow$ Fit_Transform ($D_{train,}\ D_{test}$)
II. LDA ←LDA( )
EIG_VECS ←FIT($D_X$, $D_Y$)
III. Initializing ANN
ANN← KERAS.[MODELS].SEQUENTIAL
IV. Forming hidden_Layer
ANN←ADD_LAYERS.(DENSE),UNITS←(6), ACTIVATION←RELU
V. Addition of Output_Layer
ANN←ADD_LAYERS.(DENSE),UNITS←(3), ACTIVATION←SOFTMAX
VI. Compiling_ANN
OPTIMIZER←ADAM, LOSS←{BINARY_CROSSENTROPY}
METRICS()←[ACCURACY]
VII. Fitting_ANN
Fit←($D_{train}$ , $D_{test,}$),BATCH_SIZE←32, EPOCHS←100.
VIII. Predicting result for Observation
$D_{Pred} \leftarrow EVALUATE(\ D_{Pred} > 0.5)$
IX. end procedure

# 4 Results and Discussion

Due to the fact that they are the primary cause of injuries and fatalities on the road, road accidents have become a serious issue of public health and security. Road accident injuries and passing traffic have a significant impact on individuals, families, and social structures. It weighs down nations financially because they typically have a 3% deficit in their gross domestic product (GDP).

## 4.1 Intelligent Road Accident Avoidance System

In order to build models that can predict the likelihood of car crashes, I trained ANN and other algorithms on large datasets in this work. The developed prediction model generates prediction results using real-time data gathered from various information sources such as GPS, IEEE802.11p, embedded sensors in cars, and other sources while the vehicles are in motion. A range of approaches, including LDA, dimensionality reduction algorithms, and 10 ML algorithms, are employed to increase prediction accuracy. Because the early stopping function is limited, if the validation score does not improve, the algorithm will not complete training. Table III displays the algorithm performance accuracy measurement data. True negative and false negative are represented by TN and FN, respectively, while true positive and false positive are represented by TP and FP, respectively. In eq. (5), accuracy measures the percentage of properly identified items as:

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN}x100\% \tag{5}$$

In eq. (6), precision is defined as the fraction of correct observations to the total number of projected positive observations.

$$Precision=\frac{TP}{TP+FP}x100\% \tag{6}$$

Recall, which is used as a secondary performance indicator, is calculated in eq. (7)

$$Recall=\frac{TP}{TP+FN}x100\% \tag{7}$$

Finally, the F1-score refers to the weighted average of Precision and Recall. The DT algorithm's model parameters are configured as follows: The splitting quality is evaluated using entropy, a measure of information gain. In the DT algorithm, a node is permitted to split if the split results in an impurity decrease better than or like to zero [19]. The early-stopping of tree development has been disabled. I turn off pre-sorting earlier tree splitting to hasten the process of determining the best split. The model accuracy measurements are displayed in Table 3 as Precision, Recall, and F1-score for DT. Overall, the model findings are comparable to the majority of results in other similar research, with acceptable sensitivities, fair AUC values, and mediocre specificities. The disparities are explained by changes in injury categorization criteria, independent variable sets, and the overall number of case.

## 4.2 Result Analysis

**Confusion Matrix for ANN** Severity level classified as 1, 2,3 labels. In Figure 7 it showed that total 314572 cases of fatal, serious and slight risk levels of database. True positive data showed as 7, 62, and 268583.

**Confusion Matrix for DT** Severity level classified as 1, 2,3 labels. In Figure 8 it showed that total 314572 cases of fatal, serious and slight risk levels of database. True positive data showed as 232, 7544, and 227411.

Table 4 displayed the various model performance assessment matrix parameters such as area under the curve (AUC), accuracy, F1-Score, precision, and recall. It observed that the accuracy of ANN, SGD, GB perform better than other algorithms. ANN, SGD, GB gave 85% accuracy, whereas other gave lesser accuracy percentage such as LR gave 0.854, AdaBoost 0.852, RF 0.849, NB 0.849, KNN 0.837, DT 0.749, and SVM 0.164.

The ROC curve is a plot of the true positive (TP) and false positive (FP) values. Here the graph between ANN and DT algorithms. ANN algorithm curve indicated with blue colour line where as DT algorithm line indicated with red colour which shown in Figure 9.

Table 3: Accuracy of Models along with split proportion

| **Models** | **Ratio of training set: test set** | |
|---|---|---|
| | **70:30** | **80:20** |
| DT | 0.750 | 0.749 |
| ANN | 0.855 | 0.856 |

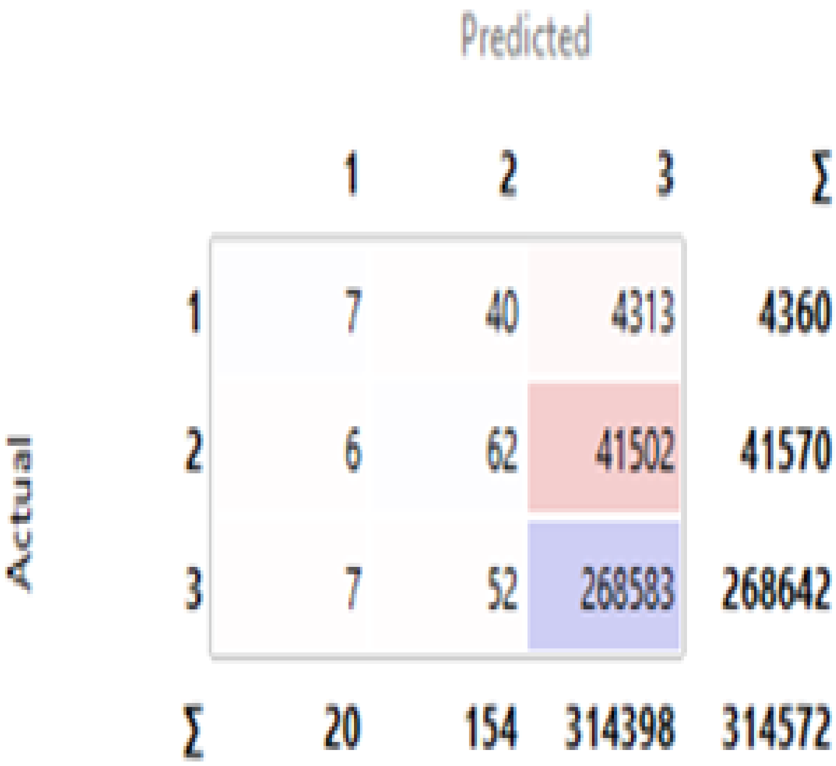


Figure 7: Confusion Matrix for ANN

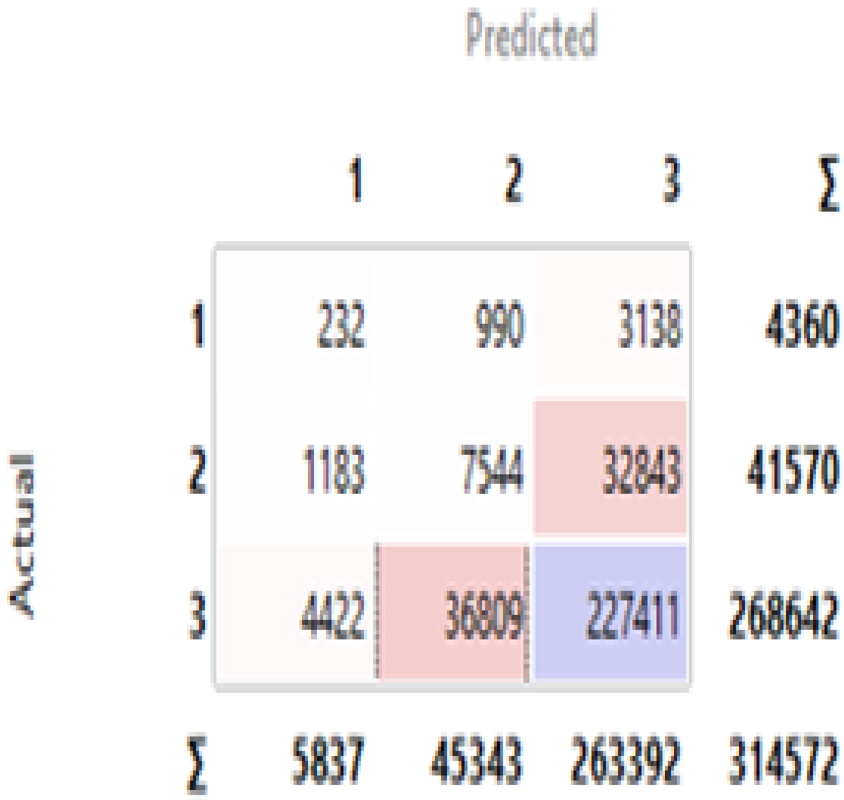


Figure 8: Confusion Matrix for DT

Table 4: Road accident severity prediction results for classification

| Model | AUC | Accuracy | F1-Score | Precision | Recall |
|---|---|---|---|---|---|
| KNN | 0.527 | 0.837 | 0.785 | 0.751 | 0.837 |
| DT | 0.511 | 0.749 | 0.754 | 0.760 | 0.749 |
| SVM | 0.535 | 0.164 | 0.246 | 0.774 | 0.164 |
| SGD | 0.500 | 0.855 | 0.788 | 0.770 | 0.855 |
| RF | 0.639 | 0.849 | 0.789 | 0.767 | 0.849 |
| ANN | 0.766 | 0.856 | 0.790 | 0.794 | 0.855 |
| NB | 0.732 | 0.849 | 0.787 | 0.778 | 0.849 |
| LR | 0.703 | 0.854 | 0.788 | 0.781 | 0.854 |
| GB | 0.765 | 0.855 | 0.788 | 0.793 | 0.855 |
| AdaBoost | 0.653 | 0.852 | 0.789 | 0.762 | 0.852 |

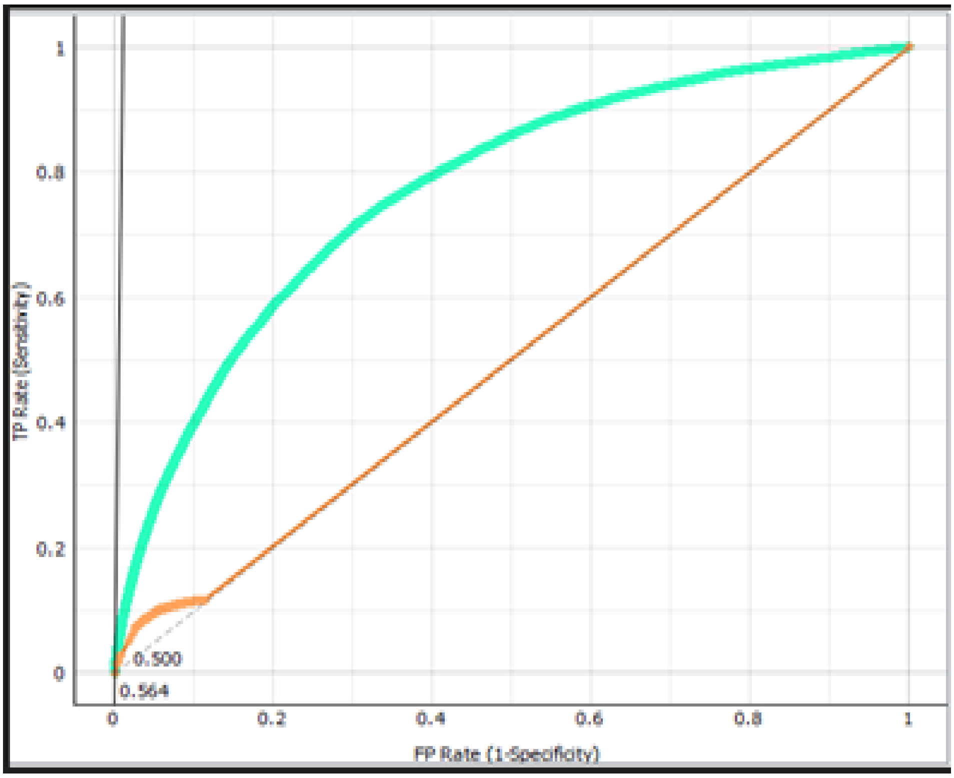


Figure 9: ROC curve for Neural network and Decision Tree

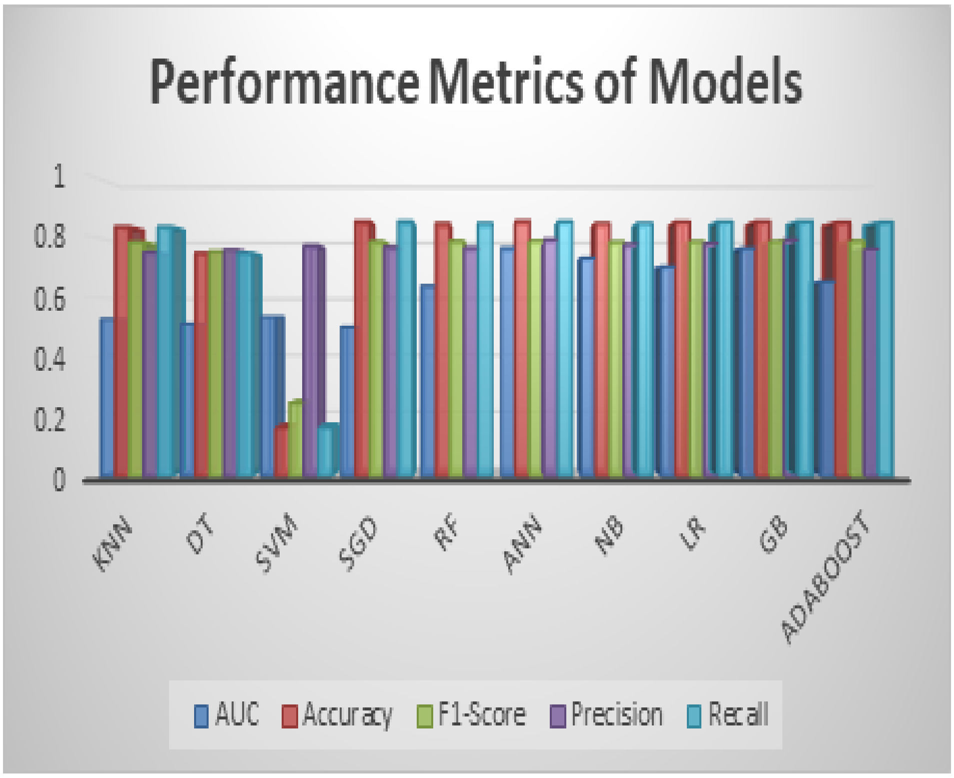


Figure 10: Accident Severity Prediction results for classification

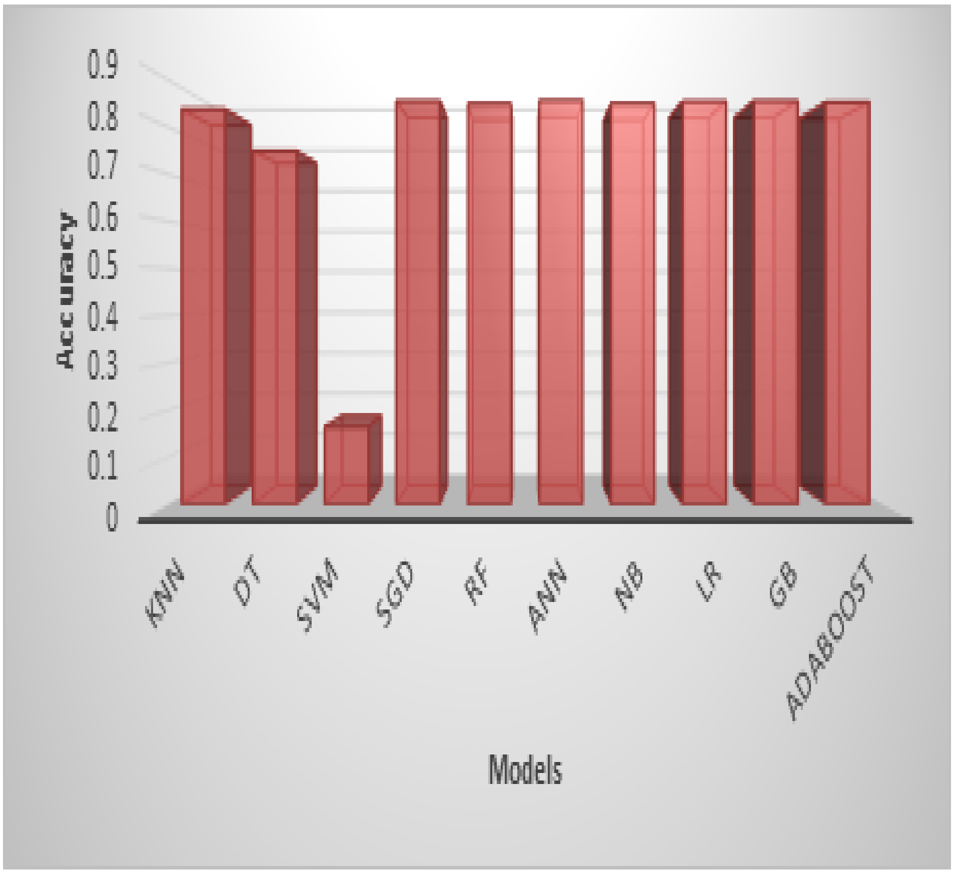


Figure 11: Accuracy of Machine Learning Models

### 4.3 Discussion

Several ML models were used to perform road accident severity prediction. Many features were identified which involves in road accident severity. To assess the severity of road accidents, the proposed model was compared to base learner models (LR, NB, DT, SVM, KNN) and an ensemble method (RF, ANN, SGD, GB, AdaBoost).

We found key qualities using the suggested model, which accounted for about half of the dataset's total accessible attributes. First, all ML models used accessible characteristics as input. In the following phase of the research, the most important qualities detected by the suggested model were used as input for all ML models.

The observed results for predicting accident severity are described in four key outcomes. First, in terms of accuracy, all ML models (ANN, SGD, and GB) fared well in predicting accident severity. ANN has the highest accuracy (0.856) when important qualities are used as input. Accuracy of SGD, and GB achieved almost similar as 0.855. SVM achieved lower accuracy as 0.164 and LR, DT, NB, KNN, AdaBoost achieve accuracy as 0.854, 0.749, 0.849, 0.837, 0.852 respectively. Accuracy outcomes of all ML models are presented in Figure 11.

Second, as seen in Figure 12, ANN beats other ML models in accuracy. However, by employing essential features, ANN achieved greater results, with a precision value of 0.794. Using all testing datasets as input, the LR achieved a precision rating of 0.781, which is less than the ANN precision score of 0.794. The RF of the Tree Ensemble model is 0.767, whereas the accuracy scores of the SGD, GB, and AdaBoost ensemble models are 0.780, 0.793, and 0.762, respectively. But precision scores of KNN, DT, SVM, NB are 0.751, 0.760, 0.774, 0.778 respectively and observed that KNN precision score is lower than compared to other algorithms.

Third, using testing data, ANN achieves a recall (sensitivity) score of 0.855, which is the maximum recall value for forecasting accident severity. Along with SGD and GB, the recall score of 0.855 is attained.

RF and NB had equal recall ratings of 0.849 utilising key features, although somewhat less than ANN, which achieved 0.856. KNN attained less recall score with 0.837 than AdaBoost with 0.852.

DT and LR achieved recall score as 0.749 and 0.854 respectively. SVM achieved lower recall score with 0.164 compared than other all algorithms. Figure 13 depicts the recall score of ML models. Fourth, examine the F1-score, which is another essential assessment statistic that combines accuracy and recall ratings.

Fourth, consider the F1-score, which is another significant assessment metric that balances accuracy and recall scores. Figure 14 observed that the F1-score of ANN (0.790) is greater than that of all other models. RF and AdaBoost achieved same F1-score as 0.789. Similarly, SGD, GB and LR achieved same F1-score with

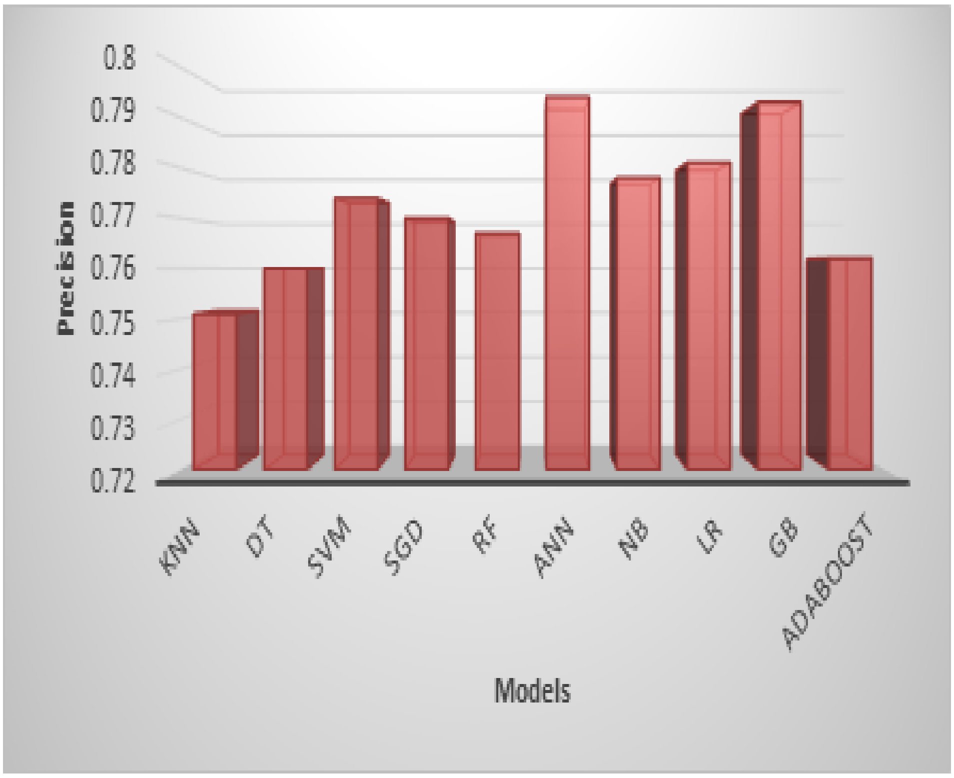


Figure 12: Precision of Machine Learning Models

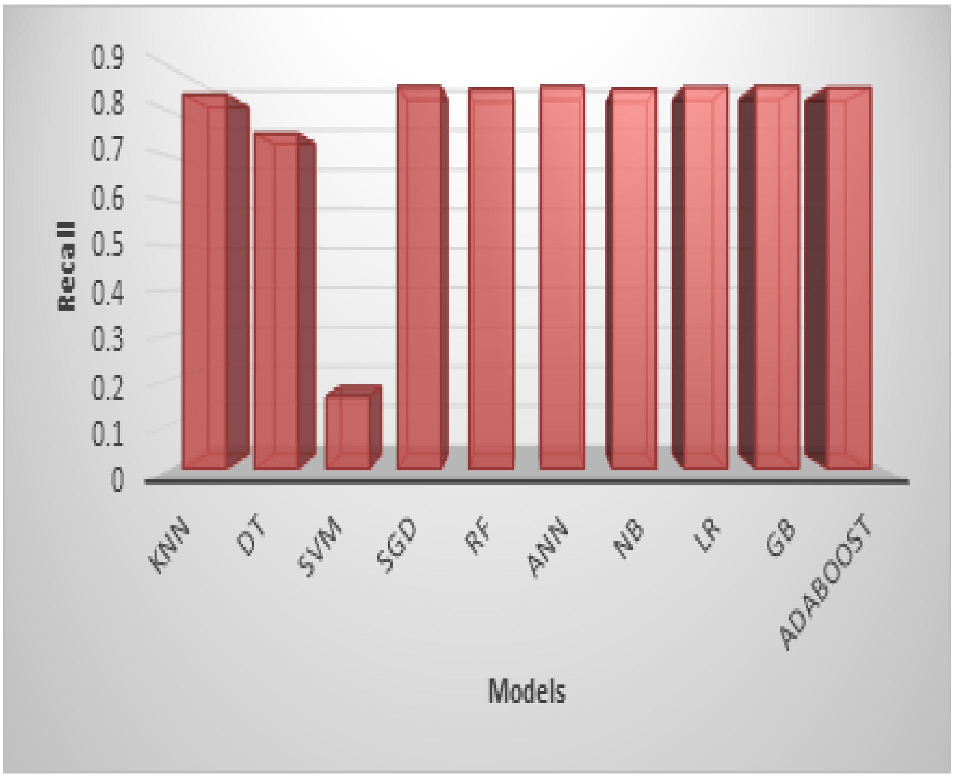


Figure 13: Recall of Machine Learning Models

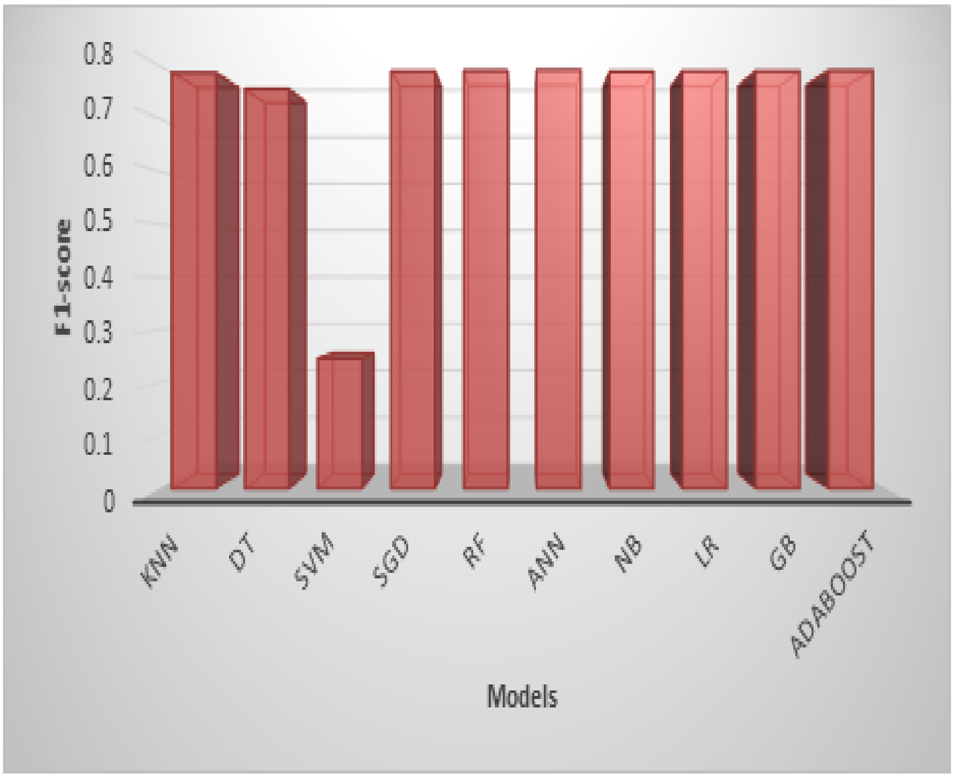


Figure 14: F1-score of Machine Learning Models

0.788. SVM achieve lower F1-score as 0.246. DT, KNN, and NB achieved F1-score as 0.754, 0.785, and 0.787 respectively.

In general, if the major focus is the entire performance of the models in forecasting the severity of road accidents, based on the summary of the results, Important characteristics that are a subset of the initial feature set are identified by models.

Using essential characteristics as input to ML models enhances accuracy, precision, recall, and F1-score. When integrated with relevant characteristics, the suggested approach not only enhances prediction performance but also lowers data gathering costs.

The road accident dataset contains 29 attributes that are used to assess the severity of the accident. Taking the 16 critical characteristics identified by the proposed model into consideration enhances the accident severity prediction procedure greatly. When all characteristics and critical features are utilised as input, the performance of the ML model is also compared. The suggested model outperformed all other models in the experiment when important characteristics were employed.

Management should focus more on the 16 most significant factors that influence accident severity.

A Receiver operating characteristic curve (ROC) is a graph that depicts a classification model's performance across all classification thresholds [3], [33]. This graph depicts two parameters: The percentage of true positives. The percentage of false positives. The degree or measure of separability is represented by AUC, whereas the probability curve is represented by ROC. It measures how well the model distinguishes between classes. The better the model predicts zero classes as zero and one class as one, the higher the AUC. The following definition of True Positive Rate (TPR), which is a synonym for recall in eq. (8) as:

$$TPR = \frac{TP}{TP+FN} \tag{8}$$

The following defines false positive rate (FPR), computed in eq. (9) to get

$$FPR = \frac{FP}{FP+TN} \tag{9}$$

At various categorization thresholds, TPR vs. FPR are displayed on a ROC curve. When the classification threshold is reduced, more objects are classified as positive, increasing the number of both false positives and true positives. "Area under the ROC Curve" is abbreviated as AUC.

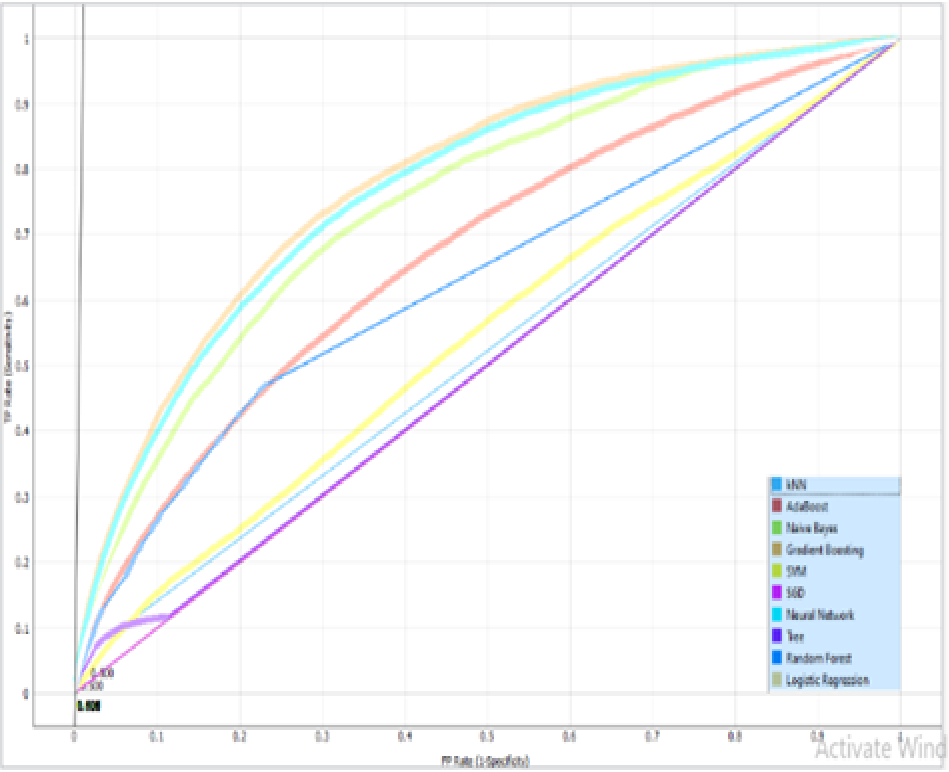


Figure 15: F1-score of Machine Learning Models

In other words, the AUC computes the two-dimensional area beneath the complete ROC curve from (0,0) to (1,1) [3].

ROC curve, drawn between TPR and FPR represented in Figure 15. It observed that dark blue colour curve represented the KNN algorithms, red colour curve represented the AdaBoost, green SGD, sky blue colour represents ANN, purple colour represents DT, royal blue colour represents RF, and grey colour represent LR. NB, SVM, and GB represents green, yellow, and brown colour respectively. Higher the value of AUC, which means higher the prediction. It found that ANN and GB both have higher value of AUC as 0.767 and 0.766 respectively.

## 5 Conclusion and Future Work

An Intelligent ANN Model estimated the severity levels of road accidents and the stages of driving automation. It also provides driving advice to reduce risk. In the pre-processing approaches, involve substitute the missing data, normalization, and standardisation, is the first step in preparing the raw data for ML. The data is then understood and useful features are chosen using techniques for data exploration and visualisation. Following that, the data were divided into two sets: training (70%) and testing (30%) and again check with 80:20 ratio. The ANN model developed for predicting the severity of road accidents got an accuracy of 0.856, while the GB and SGD models got the same accuracy of 0.855. Training on a dataset with less absent values and outliers, prediction accuracy can be further improved.

The model performance evaluation shows that the suggested ANN model outperforms the others in predicting the severity of incidents. Despite its advantages over individual models, the suggested ensemble of ML techniques improved difficulty, which will be discussed in future work. We also want to test the suggested model on a multi-domain dataset to verify its efficacy and generalizability. Because the ANN contains hyperparameters that impact its performance, it is advocated in future studies that research be done with ensemble ML and deep learning models to further enhance prediction performance.

**Declaration:**

Participation Consent and Ethical Approval:

This procedure is carried out without the involvement of people. Rights of Humans and Animals:

Animal and human rights are not being violated in any way.

**Backing:**

There is no money associated with this effort.

**Competing Interests:**

There is no potential for a conflict of interest with this project.
**Contributions to the Authorship:**
There is no evidence of authorship.
**Salutation:**
No credit is due for this creation.

# References


[1] Xie, Guo, et al. "Unmanned System Safety Decision-Making Support: Analysis and Assessment of Road Traffic Accidents." IEEE/ASME Transactions on Mechatronics 26.2 (2020): 633-644.

[2] Fu, Yuchuan, et al. "A survey of driving safety with sensing, vehicular communications, and artificial intelligence-based collision avoidance." IEEE Transactions on Intelligent Transportation Systems (2021).

[3] Chang, Yohan, et al. "Exploring Contributing Factors of Hazardous Events in Construction Zones Using Naturalistic Driving Study Data." IEEE Transactions on Intelligent Vehicles 5.3 (2020): 519-527.

[4] Hadjidimitriou, Natalia Selini, et al. "Machine learning for severity classification of accidents involving powered two wheelers." IEEE Transactions on Intelligent Transportation Systems 21.10 (2019): 4308-4317.

[5] Mansoor, Umer, et al. "Crash severity prediction using two-layer ensemble machine learning model for proactive emergency management." IEEE Access 8 (2020): 210750-210762.

[6] Nascimento, Alexandre Moreira, et al. "A systematic literature review about the impact of artificial intelligence on autonomous vehicle safety." IEEE Transactions on Intelligent Transportation Systems 21.12 (2019): 4928-4946.

[7] Shen, Xiaoyan, and Shanshan Wei. "Application of XGBoost for hazardous material road transport accident severity analysis." IEEE Access 8 (2020): 206806-206819.

[8] Zhou, Bei, et al. "Comparing factors affecting injury severity of passenger car and truck drivers." IEEE Access 8 (2020): 153849-153861.

[9] Lindow, Friedrich, et al. "Ai-based driving data analysis for behavior recognition in vehicle cabin." 2020 27th Conference of Open Innovations Association (FRUCT). IEEE, 2020.

[10] Singhal, Vivek, et al. "Artificial intelligence enabled road vehicle-train collision risk assessment framework for unmanned railway level crossings." IEEE Access 8 (2020): 113790-113806.

[11] Manzoor, Mubariz, et al. "RFCNN: traffic accident severity prediction based on decision level fusion of machine and deep learning model." IEEE Access 9 (2021): 128359-128371.

[12] Kim, J., et al. "Collision Risk Assessment Algorithm via Lane-Based Probabilistic Motion Prediction of Surrounding Vehicles......" IEEE Trans. Intell. Transp. Syst 19.9 (2018): 1-10.

[13] George, Yannis, Theofilatos Athanasios, and Pispiringos George. "Investigation of road accident severity per vehicle type." Transportation research procedia 25 (2017): 2076-2083.

[14] Halim, Zahid, et al. "Artificial intelligence techniques for driving safety and vehicle crash prediction." Artificial Intelligence Review 46.3 (2016): 351-387.

[15] Nozaki, Daichi, et al. "AI management system to prevent accidents in construction zones using 4K cameras based on 5G network." 2018 21st International Symposium on Wireless Personal Multimedia Communications (WPMC). IEEE, 2018.

[16] Tonni, Surraiya Islam, et al. "Artificial intelligence based driver vigilance system for accident prevention." 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD). IEEE, 2021.

[17] S. C. Venkateswarlu, N. U. Kumar and A. Karthik, "Speech Enhancement Using Recursive Least Square Based on Real-time adaptive filtering algorithm," 2021 6th International Conference for Convergence in Technology (I2CT), Maharashtra, India, 2021, pp. 1-4, doi: 10.1109/I2CT51068.2021.9417929.

[18] Lacatan, Luisito L., et al. "Brake-Vision: A Machine Vision-Based Inference Approach of Vehicle Braking Detection for Collision Warning Oriented System." 2021 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE). IEEE, 2021.

[19] Lin, Da-Jie, et al. "Intelligent traffic accident prediction model for Internet of Vehicles with deep learning approach." IEEE Transactions on Intelligent Transportation Systems 23.3 (2021): 2340-2349.

[20] Emu, Mahzabeen, et al. "Fatality Prediction for Motor Vehicle Collisions: Mining Big Data Using Deep Learning and Ensemble Methods." IEEE Open Journal of Intelligent Transportation Systems 3 (2022): 199-209.

[21] Thakare, Kamalakar Vijay, et al. "Object Interaction-Based Localization and Description of Road Accident Events Using Deep Learning." IEEE Transactions on Intelligent Transportation Systems (2022).

[22] D. Narsaiah, R. S. Reddy, A. Kokkula, P. A. Kumar and A. Karthik, "A Novel Full Reference-Image Quality Assessment (FR-IQA) for Adaptive Visual Perception Improvement," 2021 6th International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 2021, pp. 726-730, doi: 10.1109/ICICT50816.2021.9358610.

[23] Lin, Sung-Chiang, Chih-Jou Chen, and Tsung-Ju Lee. "A multi-label classification with hybrid label-based meta-learning method in internet of things." IEEE Access 8 (2020): 42261-42269.

[24] Karthik, A., MazherIqbal, J.L. Efficient Speech Enhancement Using Recurrent Convolution Encoder and Decoder. Wireless Pers Commun 119, 1959–1973 (2021). https://doi.org/10.1007/s11277-021-08313-6

[25] Onesimu, J. Andrew, et al. "Internet of things based intelligent accident avoidance system for adverse weather and road conditions." Journal of Reliable Intelligent Environments 7.4 (2021): 299-313.

[26] Song, Xianmin, Yaqian Sun, and Pengfei Tao. "A Dynamic Bayesian Network Model for Real-Time Risk Propagation of Secondary Rear-End Collision Accident Using Driving Risk Field." IEEE Access 10 (2022): 72429-72443.

[27] Severity Classification in Smart Transportation System." Arabian Journal for Science and Engineering (2022): 1-15.

[28] Abirami, M. S., et al. "A classification model to predict onset of smoking and drinking habits based on socio-economic and sociocultural factors." Journal of Ambient Intelligence and Humanized Computing 12.3 (2021): 4171-4179.

[29] World Health Organization, “https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries”, 19 February 2018.

[30] Ji, Ang, and David Levinson. "Injury severity prediction from two-vehicle crash mechanisms with machine learning and ensemble models." IEEE Open Journal of Intelligent Transportation Systems 1 (2020): 217-226.

[31] Samsudeen, S., and G. Senthil Kumar. "A Data-Driven Predictive Model for Speed Control in Automotive Safety Applications." IEEE Sensors Journal 22.23 (2022): 23258-23266.

[32] Umer, Muhammad, et al. "Comparison analysis of tree based and ensembled regression algorithms for traffic accident severity prediction." arXiv preprint arXiv:2010.14921 (2020).

[33] J. L. M. Iqbal and A. Karthik, "Speech Enhancement Techniques in the Digital, Analog, and Hybrid Domains," 2022 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2022, pp. 1-3, doi: 10.1109/ICCCI54379.2022.9740883.

[34] Lakshme Gowda, Monisha. Traffic Accidents Prediction Using Ensemble Machine Learning Approach. Diss. Dublin, National College of Ireland, 2020.

[35] Aldhari, Ibrahim, et al. "Severity Prediction of Highway Crashes in Saudi Arabia Using Machine Learning Techniques." Applied Sciences 13.1 (2023): 233.

[36] Department for Transport, https://data.gov.uk, Licence: Open Government Licence, Upadated: 12 October 2018.

[37] Poloju, n., & rajaram, a. (2022). Data mining techniques for patients healthcare analysis during covid-19 pandemic conditions. Journal of environmental protection and ecology, 23(5), 2105-2112.

[38] Chiranjeevi, P., & Rajaram, A. (2022). Twitter Sentiment Analysis for Environmental Weather Conditions in Recommendation of Tourism. JOURNAL OF ENVIRONMENTAL PROTECTION AND ECOLOGY, 23(5), 2113-2123.

[39] Chiranjeevi, P., & Rajaram, A. (2023). A lightweight deep learning model based recommender system by sentiment analysis. Journal of Intelligent & Fuzzy Systems, (Preprint), 1-14.



*Cite this paper as:*

Kushwaha, M.; Abirami, M.S. (2023). Intelligent Model for Avoiding Road Accidents Using Artificial Neural Network, *International Journal of Computers Communications* & *Control*, 18(5), 5317, 2023.
https://doi.org/10.15837/ijccc.2023.5.5317