IECE Transactions on Computer Science
http://dx.doi.org/10.62762/TCS.2024.766854

RESEARCH ARTICLE

Predictive Analysis for Road Safety Enhancement in Chicago County

Reshma Shaik[1], **Kislay Raj**[2], **Aditya Singh**[3] **and Teerath Kumar**[1,*]

[1] Department of Business Analytics, Dublin Business School, Dublin, Ireland
[2] School of Computing, National College of Ireland, Dublin, Ireland
[3] School of Computing, Indian Institute of Technology (BHU) Varanasi, India

Abstract

With the increasing incidents of fatal road injuries, there is an urgent need for developing effective road safety management systems. The study aims to develop predictive models based on machine learning to forecast the likelihood of road collisions depending on factors like weather, road condition, time, and driver behaviour in Chicago, USA. A machine learning approach has been applied to the crash dataset to evaluate the factors affecting the prevalence of road accidents. Python programming and the Jupyter Notebook platform have been used for performing descriptive statistics, correlation and three classification algorithms (Random Forest, KNN, Decision Tree and MLP Classification). Obtained accuracy of the KNN classifier is slightly higher than the other two classification models. The research explored insights into collision patterns related to roads, locations, and intersections. The study helps to increase road safety through targeted interventions with resource prioritisation, reducing the frequency and severity of traffic incidents by leveraging historical accident data with diverse spatial analysis techniques.

Keywords: traffic crashes, machine learning, predictive modeling, road safety, crash severity.

1 Introduction

1.1 Background of the Study

Artificial intelligence (AI) has shown a huge success in different domain such as image [1–9], audio [10–16] and many others [14, 17, 18, 28, 29]. An increase in traffic accidents is one of the many externalities brought about by the population and country expansion. Every year, road accidents result in millions of lives in addition to significant economic, social and environmental ramifications [1]. Numerous attempts such as traffic management, road clearance have been undertaken to lower the severity and frequency of the road accidents [2]. However, due to lack of proper road safety measurements, these conventional methods have not been able to properly handle road safety.

1.2 Problem Statement

There is still potential for progress in addressing this issue and more creative methods can be developed to increase the effectiveness of road safety management. Therefore, this study aims to address the road safety and collision issues in Chicago through a detailed

Academic Editor:
Jinchao Chen

Submitted: 16 October 2024
Accepted: 01 November 2024
Published: 07 December 2024

Vol. 2, **No.** 1, 2025.
10.62762/TCS.2024.766854

***Corresponding author:**
✉ Teerath Kumar
tmenghwar@staff.ncirl.ie

Citation
Shaik, R., Raj, K., Singh, A., & Kumar, T. (2025). Predictive Analysis for Road Safety Enhancement in Chicago County. *IECE Transactions on Computer Science*, 2(1), 1–9.

predictive modelling which can be built based on past road traffic and collision data.

1.3 Research Rationale

Considering the benefits of the latest technologies like ML, the research is also motivated to apply the ML models like ANN, DNN, or classification algorithms to real-life cases like road safety enhancement in Chicago. Thus, the research intends to apply the ML models for forecasting the collisions which would help in improving overall public safety and well-being.

1.4 Research Aims and Objectives

Research Aims

- Primary aim: The primary aim of the study is to identify the key factors contributing to the traffic collisions in Chicago by comprehensively analysing the crash dataset.
- Secondary aim: The secondary aim of the study is to develop predictive models that can be able to forecast the possibilities of different collisions depending on factors like weather, road condition, time, and driver behaviour.

Research Objectives

- Primary Objective: To extensively analyse the crash dataset for gaining an in-depth understanding of the factors such as weather, road conditions and driving behaviours contributing to the collisions in Chicago.
- Secondary Objectives: To develop predictive models that forecast the likelihood of different types of collisions based on factors such as weather conditions, time of day, and road conditions. Additionally, to analyze the dataset for identifying collision hotspots and uncover patterns associated with specific locations, roads, and intersections.

1.5 Research Questions

Primary Research Question

- What are the key primary factors contributing to the traffic collisions in Chicago?

Secondary Research Questions

- How well can machine learning models forecast the possibility of various collision types depending on changing conditions like the weather, state of the road, and the time of the day?
- What are the locations, roads and intersections that are related to certain collision hotspots in Chicago?

1.6 Research Hypothesis

- The machine learning models developed in this research project and the results obtained from this analysis will be extendable on other similar datasets from other parts of the world.
- Specific machine learning algorithms developed as part of this research might perform better or worse based on the nature of the data, its complexity, and the relationship between features and outcomes.
- Machine learning algorithms can uncover hidden patterns, relationships, or associations within the data that might not be immediately apparent through traditional analysis.
- Machine learning models will identify and utilise the most relevant features, thereby improving prediction accuracy and reducing overfitting.

1.7 Novelty of the research

Analysing the elements such as road conditions, and driving behaviours that influence the number of traffic accidents helps to enhance road safety. This research represents a novel approach through the application of predictive modelling for addressing the complex relationship between factors such as road condition, driver behaviour, and time of the day, contributing to road collisions.

1.8 Organisation of the Study

- The introduction explains the foundation for research while outlining the objective and reason for the study.
- The literature review reviews past publications on this issue and highlights shortcomings to be addressed.
- Research Methodology explains the models, and data collection approach, along with the method and philosophy undertaken to ethically establish the research.
- Data Analytics applies the ML models to the collected data for predictive analytics.
- The discussion summarises the findings from the models and aligns with prior research.

- The conclusion addresses the research problem concerning the findings and also provides suggestions for future study.

2 Literature Review

2.1 Trends in Traffic Accidents and Fatalities across the World

Traffic accidents have emerged as a major reason for health problems such as bone injuries, trauma, soft tissue injuries where the action and reaction of a person or an object causes personal injury as well as property damage. An estimated 1.35 million people die or become disabled as a consequence of traffic accidents each year, with vulnerable road users including pedestrians, cycles, and motorcycles accounting for a large percentage of these deaths [3]. Millions of people sustain non-fatal injuries in addition to fatal ones, which causes impairments and financial difficulties because of lost income and medical expenditures.

2.2 Factors Included in Major Traffic Accidents

The number of collisions and fatalities on the roads nowadays is one of the biggest problems facing the entire world. The primary predictors of the frequency of incidents, as identified by researchers, include inattention, reckless driving, overturn and brake failure incidents, average relative humidity and temperature, and collisions between passenger cars and pickup trucks [4]. Diverse studies showed a clear positive correlation between driving during the day and unfavourable weather conditions such as rain, fog, and slippery pavement and traffic accidents [5]. Therefore, it can be stated that both driving behavioural factors and environmental factors along with the present road condition usually influence traffic incidents, reducing road safety.

2.3 Role of Technology in Road Safety

The latest digital technologies such as AI, ML, Internet of Things (IoT), Global Positioning System (GPS), simulators, and big data are helpful in determining and offering data on factors related to road safety such as operating environment, road characteristics and road user behaviour. The impact of random variations on roadway security can be easily identified through the Smart Road Traffic Management System or SRTMS [8]. Technology-driven strategies put traffic safety first, raise driver awareness and make a major difference in preventing and reducing accidents on roads, which raise the overall road safety standards.

2.4 Predictive Modelling for Enhancing Road Safety

On highways, traffic accidents continue to be the primary reason for fatalities, serious injuries and major disruptions. Diverse research studies also developed a dataset to identify potentially risky routes in response to an increase in traffic accidents, which disproportionately harm women [7]. Establishing a high-precision model that presents the likelihood of every category of future accidents may be achieved by modelling the magnitude of accidents utilising the most effective factors like behavioural and environmental factors. Therefore, it can be stated that predictive modelling using advanced algorithms offer promising solutions for enhancing road safety, however, their effectiveness is highly dependent on the data and models used.

2.5 Literature gap

Even though the analysis of prior research showcases a comprehensive overview of the present research problem, there is still room for improvement. For more precise accident prediction and prevention, additional research is needed to create predictive models that take into account real-time data [6, 9]. Therefore, this research seeks to consider the diverse factors influencing road accidents in Chicago, USA where the majority of the traffic incidents are reported, with the application of latest predictive modelling algorithms like Random Forest.

3 Methodology

3.1 Proposed Methodological Architecture

The above methodological architecture (as shown in Figure 1) mainly follows a structured approach emphasising diverse stages such as data loading, importing necessary libraries, and data exploration involving analysis of its shape, present columns, and dataset information. Data cleaning and preprocessing contain handling null values, dropping irrelevant columns, encoding categorical features, and grouping target variable total injuries. Data visualisations then help to understand the patterns and trends of the features enlisted in the data. After that, data splitting helps to train and test the predictive model for accuracy prediction in this context. Predictive models such as Random Forest Classifier, K-Nearest Neighbour Classifier, Decision Tree Classifier and MLP Classification are performed with significant metrics of accuracy percentage, classification report. The Comparison based on the accuracy score explores the next-fitted model to predict the total injuries

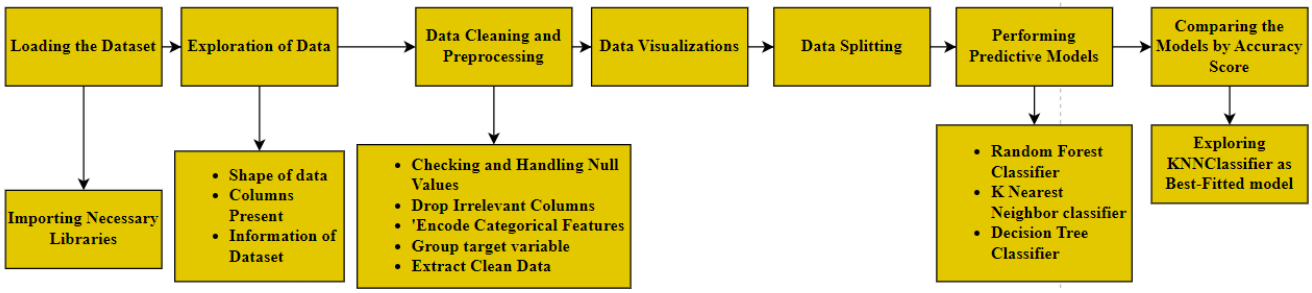

Figure 1. Flow Chart of the process of developing the artefact.

adequately.

3.2 Data Collection

This dataset enables a thorough exploration of factors impacting traffic accidents, providing informed decision-making with accident prevention strategies. Initially, a case study of Montgomery County was selected but after extensive research it was found that the Chicago county of the US has the highest number of road crashes as well as traffic related problems as compared to Montgomery County. This has helped to gather a dataset which helps to understand all the factors leading to traffic crashes leading to developing the models with proper results.

3.3 Machine Learning Models

Machine learning models like Random Forest, K-Nearest Neighbour Classifier,Decision Tree Classifier and MLP Classifier are suitable for analysing crash data due to their capability to handle complicated patterns and nonlinear relationships within the data. Thus, these effective machine learning models are employed in this context to analyse traffic crash data adequately.

3.4 Chapter Summary

With approaching a positivist belief, deductive technique and quantitative analysis, multiple ML models have been applied to the Chicago crash data. This methodology chapter has thus demonstrated the systematic approach taken for this research to properly develop the prediction models through ML algorithms.

4 Data Analysis and Findings

4.1 Introduction

Data analysis is the process of evaluating data for the fulfilment of the research aim and objectives. In this context, a machine learning approach of data analysis has been performed to evaluate the factors affecting road crashes in Chicago city through the application of descriptive (summary statistics, visualisation and correlation) and predictive analytics (classification algorithms).

4.2 Dataset Exploration

```
data.shape
```

```
(746498, 49)
```

```
data.columns
```

```
Index(['CRASH_RECORD_ID', 'RD_NO', 'CRASH_DATE_EST_I', 'CRASH_DATE',
       'POSTED_SPEED_LIMIT', 'TRAFFIC_CONTROL_DEVICE', 'DEVICE_CONDITION',
       'WEATHER_CONDITION', 'LIGHTING_CONDITION', 'FIRST_CRASH_TYPE',
       'TRAFFICWAY_TYPE', 'LANE_CNT', 'ALIGNMENT', 'ROADWAY_SURFACE_COND',
       'ROAD_DEFECT', 'REPORT_TYPE', 'CRASH_TYPE', 'INTERSECTION_RELATED_I',
       'NOT_RIGHT_OF_WAY_I', 'HIT_AND_RUN_I', 'DAMAGE', 'DATE_POLICE_NOTIFIED',
       'PRIM_CONTRIBUTORY_CAUSE', 'SEC_CONTRIBUTORY_CAUSE', 'STREET_NO',
       'STREET_DIRECTION', 'STREET_NAME', 'BEAT_OF_OCCURRENCE',
       'PHOTOS_TAKEN_I', 'STATEMENTS_TAKEN_I', 'DOORING_I', 'WORK_ZONE_I',
       'WORK_ZONE_TYPE', 'WORKERS_PRESENT_I', 'NUM_UNITS',
       'MOST_SEVERE_INJURY', 'INJURIES_TOTAL', 'INJURIES_FATAL',
       'INJURIES_INCAPACITATING', 'INJURIES_NON_INCAPACITATING',
       'INJURIES_REPORTED_NOT_EVIDENT', 'INJURIES_NO_INDICATION',
       'INJURIES_UNKNOWN', 'CRASH_HOUR', 'CRASH_DAY_OF_WEEK', 'CRASH_MONTH',
       'LATITUDE', 'LONGITUDE', 'LOCATION'],
      dtype='object')
```

Figure 2. Shape and column of the dataset.

The shape of the dataset has been checked using the 'shape' function in Python, from which the observed shape of the dataset is (746498,49), indicating the dataset has 746498 observations and 49 features (Refer to Figure 2).

4.3 Data Preprocessing

Null values in the dataset have been checked using the $isnull().sum()$ function in Python. As a result, the columns with missing values have been dropped from the dataset using the 'data.drop' function in Python, reducing data redundancy in the dataset (Refer to Figure 3). Mannualy removed columns are shown in Figure 4.

Figure 5 demonstrates that the variables 'INJURIES TOTAL' and 'LOCATION' contain 1619 and 4988 missing values. Due to this, these two columns ('INJURIES TOTAL' and 'LOCATION') have been dropped from the dataset, helping in the improvement of the structural integrity of the dataset.

```
#Checking Missing Values
data.isnull().sum()
```

```
CRASH_RECORD_ID                0
RD_NO                       4307
CRASH_DATE_EST_I          690109
CRASH_DATE                     0
POSTED_SPEED_LIMIT             0
TRAFFIC_CONTROL_DEVICE         0
DEVICE_CONDITION               0
WEATHER_CONDITION              0
LIGHTING_CONDITION             0
FIRST_CRASH_TYPE               0
TRAFFICWAY_TYPE                0
LANE_CNT                  547494
ALIGNMENT                      0
ROADWAY_SURFACE_COND           0
ROAD_DEFECT                    0
REPORT_TYPE                21222
CRASH_TYPE                     0
INTERSECTION_RELATED_I    575368
NOT_RIGHT_OF_WAY_I        711724
HIT_AND_RUN_I             513706
DAMAGE                         0
DATE_POLICE_NOTIFIED           0
PRIM_CONTRIBUTORY_CAUSE        0
SEC_CONTRIBUTORY_CAUSE         0
STREET_NO                      0
STREET_DIRECTION               4
STREET_NAME                    1
BEAT_OF_OCCURRENCE             5
PHOTOS_TAKEN_I            737007
STATEMENTS_TAKEN_I        730402
DOORING_I                 744212
WORK_ZONE_I               742175
WORK_ZONE_TYPE            743128
WORKERS_PRESENT_I         745383
NUM_UNITS                      0
MOST_SEVERE_INJURY          1630
INJURIES_TOTAL              1619
```

Figure 3. Sample of Null values in the dataset.

```
# Dropping columns with too many missing values
columns_to_drop = [
    'CRASH_RECORD_ID', 'RD_NO', 'CRASH_DATE_EST_I', 'CRASH_DATE', 'LANE_CNT',
    'ALIGNMENT', 'REPORT_TYPE', 'INTERSECTION_RELATED_I', 'NOT_RIGHT_OF_WAY_I',
    'HIT_AND_RUN_I', 'DATE_POLICE_NOTIFIED', 'STREET_NO', 'STREET_DIRECTION',
    'STREET_NAME', 'PHOTOS_TAKEN_I', 'STATEMENTS_TAKEN_I', 'DOORING_I',
    'WORK_ZONE_I', 'WORK_ZONE_TYPE', 'WORKERS_PRESENT_I', 'NUM_UNITS',
    'INJURIES_FATAL', 'INJURIES_INCAPACITATING', 'INJURIES_NON_INCAPACITATING',
    'INJURIES_REPORTED_NOT_EVIDENT', 'INJURIES_NO_INDICATION', 'INJURIES_UNKNOWN',
    'LATITUDE', 'LONGITUDE', 'MOST_SEVERE_INJURY', 'PRIM_CONTRIBUTORY_CAUSE', 'SEC_CONTRIBUTORY_CAUSE',
]
data = data.drop(columns=columns_to_drop)
```

Figure 4. Dropping the columns that contain a high number of missing values.

```
#Checking Missing Values
data.isnull().sum()
```

```
POSTED_SPEED_LIMIT          0
TRAFFIC_CONTROL_DEVICE      0
DEVICE_CONDITION            0
WEATHER_CONDITION           0
LIGHTING_CONDITION          0
FIRST_CRASH_TYPE            0
TRAFFICWAY_TYPE             0
ROADWAY_SURFACE_COND        0
ROAD_DEFECT                 0
CRASH_TYPE                  0
DAMAGE                      0
BEAT_OF_OCCURRENCE          5
INJURIES_TOTAL           1619
CRASH_HOUR                  0
CRASH_DAY_OF_WEEK           0
CRASH_MONTH                 0
LOCATION                 4908
dtype: int64
```

```
# Drop rows with missing values in the specified columns
columns_to_check = ['BEAT_OF_OCCURRENCE','INJURIES_TOTAL']
data = data.dropna(subset=columns_to_check)
data.head()
```

Figure 5. Checking missing values in the dataset after the removal of columns that contain a high number of missing values.

```
from sklearn.model_selection import train_test_split

# Split the data into features (X) and target variable (y)
X = data.drop(columns=['INJURY_LEVEL', 'CRASH_DAY_OF_WEEK', 'CRASH_HOUR', 'CRASH_MONTH'])
y = data['INJURY_LEVEL']

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Display the shapes of the training and testing sets
print("Training set shape - X:", X_train.shape, "y:", y_train.shape)
print("Testing set shape - X:", X_test.shape, "y:", y_test.shape)
```

```
Training set shape - X: (595899, 13) y: (595899,)
Testing set shape - X: (148975, 13) y: (148975,)
```

Figure 6. Data Splitting.

The data is split into training and testing sets by using an 80-20 ratio, as shown in Figure 6. This specific approach helps prevent overfitting by allowing the models to eventually generalise to unseen data, thus offering a more reliable assessment of their predictive ability (Refer to Figure 6).

4.4 Exploratory Data Analysis and Data Visualisations

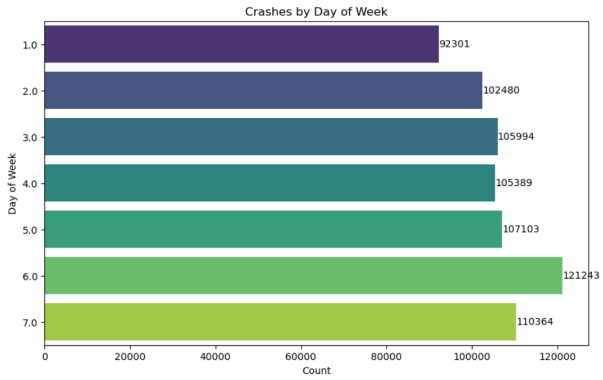

Figure 7. Distribution of crashes by day of the week.

Figure 7 demonstrates the distribution of crashes across the day of the week, from which it can be observed that the number of crashes was comparatively higher on Saturday (121243) and Sunday (110364). On the other hand, the prevalence of crashes was comparatively low (92301) on Monday in Chicago.

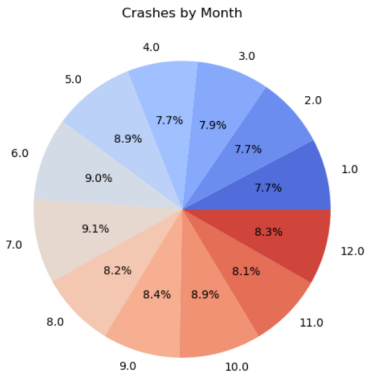

Figure 8. Distribution of crashes by month.

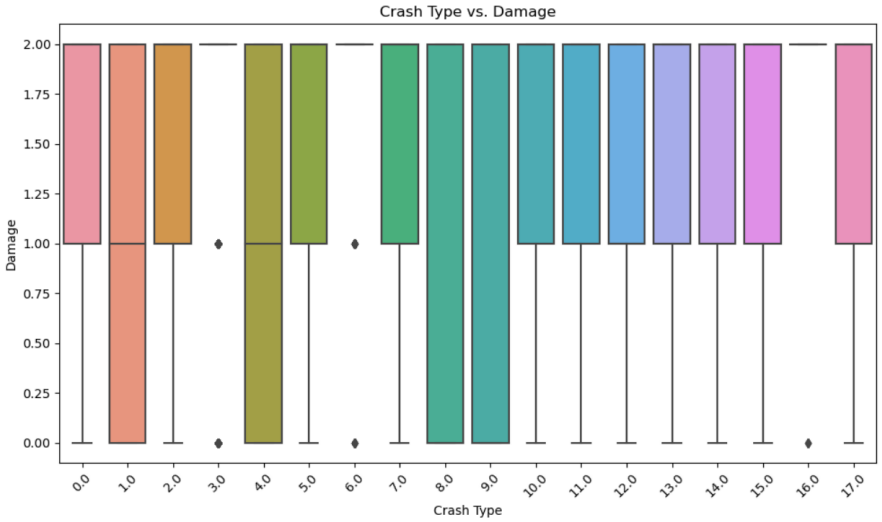

Figure 9. Distribution of Crash type by Damage.

Figure 8 shows the distribution of crashes in Chicago in each of the 12 months of the year. From the above Figure, it can be inferred that the prevalence of crashes was nearly the same (approximately (8percent-9 percent) across the 12 months of the year.

Figure 9 demonstrates the distribution of crash types by damage level, from which it can be observed that damage levels were highest for crash types 8 and 9. On the other hand, the damage levels for crash types 0, 7, 15 and 17 (less severe crashes) were considerably low. This leads to the possible indication that Crash types 8 and 9 may involve scenarios such as head-on collisions or collisions with fixed objects like trees or barriers, which fundamentally result in more extensive damage due to the high-impact forces involved.

```
RandomForestClassifier
RandomForestClassifier(random_state=42)
```

```
# Predictions on the testing set
y_pred_rf = rf_classifier.predict(X_test)
```

```
# Calculate accuracy score
accuracy = accuracy_score(y_test, y_pred_rf)
print("Accuracy Score: {:.2f}%".format(accuracy * 100))
```

```
Accuracy Score: 99.82%
```

```
# Classification report
print("\nClassification Report:")
print(classification_report(y_test, y_pred_rf))
```

```
Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00    148735
           1       0.05      0.00      0.01       232
           2       0.00      0.00      0.00         8

    accuracy                           1.00    148975
   macro avg       0.35      0.33      0.34    148975
weighted avg       1.00      1.00      1.00    148975
```

Figure 10. Evaluation of Random Forest Classifier.

4.5 Predictive Models

4.5.1 Random Forest Classifier

The random forest classifier is trained on the training set and accomplished an accuracy score of 99.82 percentage on the testing set, signifying high performance, as shown in Figure 10. Moreover, upon close inspection using the classification report, it is observed that the model struggles with minority classes.

4.5.2 KNN Classifier

```
KNeighborsClassifier
KNeighborsClassifier()
```

```
# Predict on the testing data
y_pred_knn = knn_classifier.predict(X_test)
```

```
# Calculate accuracy score
accuracy = accuracy_score(y_test, y_pred_knn)
print("Accuracy Score: {:.2f}%".format(accuracy * 100))
```

```
Accuracy Score: 99.84%
```

```
# Classification report
print("\nClassification Report:")
print(classification_report(y_test, y_pred_knn))
```

```
Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00    148735
           1       0.00      0.00      0.00       232
           2       0.00      0.00      0.00         8

    accuracy                           1.00    148975
   macro avg       0.33      0.33      0.33    148975
weighted avg       1.00      1.00      1.00    148975
```

Figure 11. Evaluation of KNN Classifier.

The KNN classifier accomplishes a commendable accuracy score of 99.84 percent on the testing crash data. However, its evaluation performance on minority classes (1 and 2) is insignificant as shown by the classification report in Figure 11.

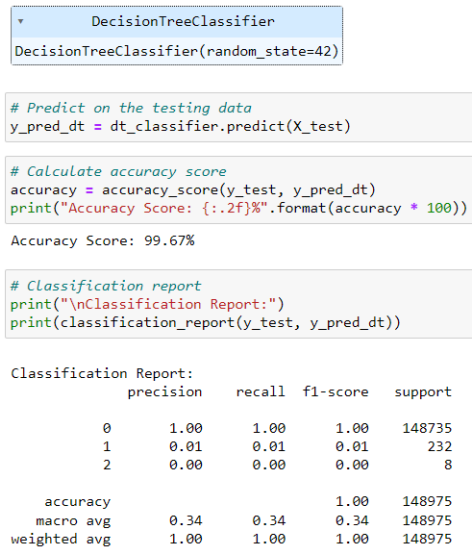

Figure 12. Evaluation Performance of Decision Tree Classifier.

4.5.3 *Decision Tree Classifier*

The Decision Tree classifier illustrates a robust accuracy of 99.67 percent, as evidenced by the classification report in Figure 12. However, it also struggles similarly to the previous classification models in accurately predicting minority classes with low precision and recall scores for classes 1 and 2.

4.5.4 *MLP Classifier*

Figure 13. Classification report for MLP classifier.

The accuracy obtained from the MLP classifier is 0.9984, indicating that 99.84 percent of the instances were accurately captured by the MLP model (Refer to Figure 13). This highlights the significantly high accuracy of the model for the prediction of road crashes in Chicago.

4.6 Comparison Between the Model Accuracy

Table 1, represents the classification accuracy of the three classification models, from which the obtained accuracy for the KNN model is slightly higher (99.84 percent) than the other two models MLP classifier (99.84 percent), Random Forest (99.82 percent) and Decision tree classifier model (99.67 percent).

Table 1. Classifier Accuracy Scores.

Classifier	Accuracy Score (%)
Random Forest	99.82
KNN	99.84
Decision Tree	99.67
MLP Classifier	99.84

5 Discussion

5.1 Summary of findings

Upon splitting the data into training and testing sets with an 80-20 ratio, classification models are trained as well as evaluated, involving Random Forest, KNN, and Decision Tree classifiers. These classification models mainly exhibit high overall accuracy, accomplishing utmost accuracy, they struggle with minority classes (1 and 2) as evidenced by low precision and recall scores in the classification reports. This highlights potential class imbalance issues, where the predictive models inadequately favour the majority class (0) and struggle to accurately predict instances from minority classes.

5.2 Discussion of the findings with respect to prior research

A comprehensive understanding of these factors impacting road traffic incidents can easily be gained by analysing the crash dataset. This analysis identifies risk factors and evaluates targeted interventions to enhance road safety. Primary Objective has been discussed through evidenced-based interventions and analysing of the crash dataset to gain an in-depth understanding of each significant feature. The entire analysis explores patterns and correlations within the data, enabling the identification of each key predictor with the evaluation of these predictive models. Thus, Secondary Objective has been discussed through employing the classification models for predicting the likelihood of distinct types of collisions based on the predictors enlisted in the data. In addition to that, spatial analysis techniques can also be employed to recognise clusters of accidents to assess their proximity to diverse road features with infrastructures. Thus, Secondary Objective has been discussed by performing analysis of the data to leverage data-driven insights and address collision hotspots adequately.

6 Conclusion and Future Work

Based on the above discussion it can be stated that there exists a significant relationship between fatal road traffic incidents and variables such as type of road, impaired visibility, accident location, weather, and timing of the incident. Findings from past literature through the utilisation of different crash datasets have emphasised the existence of a significant correlation between the fatality of road accidents and factors such as type of road, impaired visibility of the drivers, bad weather conditions and timing of accidents.Establishing a high-precision approach that presents the likelihood of every category of future accidents can be achieved by modelling the magnitude of accidents utilising the most effective factors like behavioural and environmental factors. This model can then be used to help authorities choose remedial actions related to identification of accident-prone zones or areas.

Future studies should incorporate a wider range of quantitative data to enhance the depth and generalizability of the findings. Additionally, it would be beneficial to explore driving behavior and the effectiveness of the traffic control system, as these factors could provide further insights. Furthermore, addressing the model's limitations and discussing potential strategies for improving performance on minority classes will be crucial in future research.

Conflicts of Interest

The authors declare no conflicts of interest.

Funding

This work was supported without any funding.

References

[1] Kumar, T., Mileo, A., & Bendechache, M. (2024, June). Keeporiginalaugment: Single image-based better information-preserving data augmentation approach. In *IFIP International Conference on Artificial Intelligence Applications and Innovations* (pp. 27-40). Cham: Springer Nature Switzerland.

[2] Roy, A. M., Bhaduri, J., Kumar, T., & Raj, K. (2022). A computer vision-based object localization model for endangered wildlife detection. *Ecological Economics*, Forthcoming.[CrossRef]

[3] Kumar, T., Brennan, R., Mileo, A., & Bendechache, M. (2024). Image data augmentation approaches: A comprehensive survey and future directions. *IEEE Access*.[CrossRef]

[4] Kumar, T., Mileo, A., Brennan, R., & Bendechache, M. (2023). RSMDA: Random Slices Mixing Data Augmentation. *Applied Sciences*, 13(3), 1711.[CrossRef]

[5] Chandio, A., Gui, G., Kumar, T., Ullah, I., Ranjbarzadeh, R., Roy, A. M., ... & Shen, Y. (2022). Precise single-stage detector. *arXiv preprint arXiv:2210.04252*.[CrossRef]

[6] Kumar, T., Turab, M., Raj, K., Mileo, A., Brennan, R. & Bendechache, M. (2023). Advanced Data Augmentation Approaches: A Comprehensive Survey and Future directions. *ArXiv Preprint ArXiv:2301.02830*.

[7] Kumar, T., Park, J., Ali, M. S., Uddin, A. F. M., & Bae, S. H. (2021). Class specific autoencoders enhance sample diversity. *Journal Of Broadcast Engineering*, 26(7), 844-854.[CrossRef]

[8] Aleem, S., Kumar, T., Little, S., Bendechache, M., Brennan, R., & McGuinness, K. (2022). Random data augmentation based enhancement: a generalized enhancement approach for medical datasets. *arXiv preprint arXiv:2210.00824*.[CrossRef]

[9] Kumar, T., Park, J., Ali, M. S., Uddin, A. S., Ko, J. H., & Bae, S. H. (2021). Binary-classifiers-enabled filters for semi-supervised learning. *IEEE Access*, 9, 167663-167673.[CrossRef]

[10] Chandio, A., Shen, Y., Bendechache, M., Inayat, I., & Kumar, T. (2021). AUDD: audio Urdu digits dataset for automatic audio Urdu digit recognition. *Applied Sciences*, 11(19), 8842. [CrossRef]

[11] Turab, M., Kumar, T., Bendechache, M., & Saber, T. (2022). Investigating multi-feature selection and ensembling for audio classification. *arXiv preprint arXiv:2206.07511*. [CrossRef]

[12] Raj, K., Singh, A., Mandal, A., Kumar, T., & Roy, A. M. (2023). Understanding EEG signals for subject-wise definition of armoni activities. *arXiv preprint arXiv:2301.00948*.[CrossRef]

[13] Kumar, T., Park, J., & Bae, S. H. (2020). Intra-Class Random Erasing (ICRE) augmentation for audio classification. In *Proceedings Of The Korean Society Of Broadcast Engineers Conference* (pp. 244-247). The Korean Institute of Broadcast and Media Engineers.

[14] Park, J., Kumar, T., & Bae, S. H. (2020). Search for optimal data augmentation policy for environmental sound classification with deep neural networks. *Journal Of Broadcast Engineering*, 25(6), 854-860. [CrossRef]

[15] Park, J., Kumar, T., & Bae, S. H. (2020). Search of an optimal sound augmentation policy for environmental sound classification with deep neural networks. In *Proceedings Of The Korean Society Of Broadcast Engineers Conference* (pp. 18-21). The Korean Institute of Broadcast and Media Engineers.

[16] Kumar, T., Turab, M., Mileo, A., Bendechache, M., & Saber, T. (2023). AudRandAug: Random Image Augmentations for Audio Classification. *arXiv preprint arXiv:2309.04762*. [CrossRef]

[17] Singh, A., Raj, K., Meghwar, T., & Roy, A. M. (2024). Efficient Paddy Grain Quality Assessment Approach Utilizing Affordable Sensors. *AI*, 5(2), 686-703.[CrossRef]

[18] Khan, W., Kumar, T., Zhang, C., Raj, K., Roy, A. M., & Luo, B. (2023). SQL and NoSQL database software architecture performance analysis and assessments—a systematic literature review. *Big Data and Cognitive Computing*, 7(2), 97. [CrossRef]

[19] Silva, P. B., Andrade, M., & Ferreira, S. (2020). Machine learning applied to road safety modeling: A systematic literature review. *Journal of traffic and transportation engineering* (*English edition*), 7(6), 775-790. [CrossRef]

[20] Gebresenbet, R. F., & Aliyu, A. D. (2019). Injury severity level and associated factors among road traffic accident victims attending emergency department of Tirunesh Beijing Hospital, Addis Ababa, Ethiopia: a cross sectional hospital-based study. *PLoS One*, 14(9), e0222793. [CrossRef]

[21] Ahmed, S. K., Mohammed, M. G., Abdulqadir, S. O., El-Kader, R. G. A., El-Shall, N. A., Chandran, D., ... & Dhama, K. (2023). Road traffic accidental injuries and deaths: A neglected global health issue. *Health science reports*, 6(5), e1240.[CrossRef]

[22] Behzadi Goodari, M., Sharifi, H., Dehesh, P., Mosleh-Shirazi, M. A., & Dehesh, T. (2023). Factors affecting the number of road traffic accidents in Kerman province, southeastern Iran (2015–2021). *Scientific reports*, 13(1), 6662.

[23] Lin, D. J., Yang, J. R., Liu, H. H., Chiang, H. S., & Wang, L. Y. (2022). Analysis of environmental factors on intersection accidents. *Sustainability*, 14(3), 1764. [CrossRef]

[24] Nižetić, S., Šolić, P., Gonzalez-De, D. L. D. I., & Patrono, L. (2020). Internet of Things (IoT): Opportunities, issues and challenges towards a smart and sustainable future. *Journal of cleaner production*, 274, 122877. [CrossRef]

[25] Satla, S. P., Sadanandam, M., & Suvarna, B. (2020). Dangerous Prediction in Roads by Using Machine Learning Models. *Ingénierie des Systèmes d'Information*, 25(5).

[26] Sharma, A., Awasthi, Y., & Kumar, S. (2020, October). The role of blockchain, AI and IoT for smart road traffic management system. In *2020 IEEE India Council International Subsections Conference* (INDISCON) (pp. 289-296). IEEE. [CrossRef]

[27] Tonhauser, M., & Ristvej, J. (2021). Implementation of new technologies to improve safety of road transport. *Transportation research procedia*, 55, 1599-1604. [CrossRef]

[28] Kumar, T., Bhujbal, R., Raj, K., & Roy, A. M. (2024). Navigating Complexity: A Tailored Question-Answering Approach for PDFs in Finance, Bio-Medicine, and Science.

[29] Barua, M., Kumar, T., Raj, K., & Roy, A. M. (2024). Comparative Analysis of Deep Learning Models for Stock Price Prediction in the Indian Market.

Reshma Shaik graduated with a Master's degree in Data Analytics from Dublin Business School in May 2024, achieving a 2.1 Honors. Her thesis, titled "Predictive Analysis for Road Safety Enhancement in Chicago County," utilized machine learning models to predict road collisions based on various factors, with the K-Nearest Neighbors (KNN) classifier achieving the highest accuracy. This research was published on Preprints.org, providing insights for targeted safety interventions.
Reshma previously earned a Bachelor of Technology in Electrical and Electronics Engineering from Vignan's Lara Institute of Technology and Science. With a strong passion for data-driven solutions, she aims to contribute to advancements in safety and operational efficiency. (Email: 20007371@mydbs.ie)

Kislay Raj is a highly accomplished Associate Professor at the National College of Ireland. He holds a Master's in Data Analytics with First Class Distinction from Ireland and has accrued valuable industry experience in diverse roles, including data scientist, senior lead data analyst, marketing analyst, and software engineer.
His cutting-edge research focuses on Neuro-Symbolic Artificial Intelligence, an innovative field that seeks to bridge the gap between deep learning and reasoning for explainable and trustworthy decision support. Kislay's research achievements are underscored by his publication record in top-tier academic journals, including papers in the fields of computer vision, explainable AI, machine learning, and robotics. His broad research interests encompass trustworthy AI, deep learning, explainable machine learning, and computer vision, along with their practical applications to real-world systems. (Email: kraj@staff.ncirl.ie)

Aditya Singh is an Assistant Professor at the School of Computing, Indian Institute of Technology (BHU) Varanasi, India. He completed his Ph.D. at the Center of Intelligent Robotics, Indian Institute of Information Technology Allahabad. Aditya earned his M. Tech. in Robotics from IIIT Allahabad in 2018 and his B. Tech. in Mechanical Engineering from Dr. A. P. J. Abdul Kalam Technical University, Lucknow, in 2016.
He has been actively involved with SAE and the IEEE-RAS UP Chapter and has published various papers in the fields of robotics and computer vision. His research interests include mobile robotics, deep learning, machine learning, and their applications to real-world systems. (Email: rsi2018003@iiit.ac.in)

Teerath Kumar is a Dissertation Supervisor in the Department of Business Analytics at Dublin Business School and serves as Associate Faculty at the National College of Ireland. With expertise in data analytics and business intelligence, he guides students in their research endeavors, helping them navigate complex analytical frameworks and methodologies. Teerath is committed to fostering a collaborative learning environment and enhancing students' understanding of the practical applications of business analytics in today's data-driven landscape. (Email: tmenghwar@staff.ncirl.ie)