Road Accident Prediction And Classification

**PROJECT SYNOPSIS**

**OF MAJOR PROJECT**

**BACHELOR OF TECHNOLOGY**

**CSE**

**SUBMITTED BY: PCSE25-69**

**August 2023**

****

**KIET Group of Institutions, Delhi-NCR,**

**Ghaziabad (UP)**

**Department of Computer Science and Engineering**

**KIET Group of Institutions, Delhi-NCR,** **Ghaziabad (UP)**

**Department of Computer Science and Engineering**

|  |  |  |  |
| --- | --- | --- | --- |
| **Name of Student** | **Branch** | **University Roll no.** | **Class Roll No.** |
| * **Harshit Goel** | **CSE** | **2100290310065** | **66** |
| * **Mohit Saini** | **CSE** | **2100290100100** | **34** |
| * **Harsheet** | **CSE** | **2100290100070** | **4** |

**Table of Contents:**

|  |  |
| --- | --- |
| **Content** | **Page No.** |
| **1. Introduction** | **03** |
| **2. Rationale** | **04** |
| **3. Literature Review** | **05** |
| **4. Feasibility Study** | **09** |
| **5. Methodology/Planning of work** | **10** |
| **6. Facilities required for proposed work** | **11** |
| **7. Expected outcomes** | **12** |
| **8. References** | **12** |

**Introduction**

Road traffic accidents represent a critical and complex challenge globally, resulting in approximately **1.2 million deaths and 50 million injuries every year**. These accidents incur vast **economic, social, and emotional costs**, with an estimated **$43 billion annually** lost due to traffic crashes.

Traditional methods for predicting road accidents primarily use **linear statistical models** and often fail to capture the **multi-factorial, non-linear relationships** that characterize real-world traffic conditions. These approaches also lack the capacity to deliver **real-time, actionable insights** for traffic management.

This project proposes a **machine learning-based prediction system** that analyzes multiple factors—such as **environmental conditions (weather, lighting)**, **road characteristics (surface, traffic density)**, and **driver demographics (age, gender)**—to accurately classify road accidents based on severity levels: **Fatal, Serious, or Slight**. The system leverages **classification algorithms**, cloud-based deployment, and interactive visualizations to enable **proactive decision-making** by traffic authorities.

**Problem Statement**

The complexity of road accident causation involves multiple interconnected factors including driver behavior, environmental conditions, vehicle characteristics, and infrastructure elements. Traditional prediction approaches fail to adequately model these relationships, resulting in:

1. Insufficient accuracy in severity prediction
2. Inability to provide real-time risk assessment
3. Limited integration with preventive action systems
4. Poor scalability for large-scale deployment

**Project Objectives**

This project aims to:

1. Develop Advanced Prediction Models: Create machine learning algorithms capable of accurately predicting accident severity based on multiple input parameters
2. Build Real-time System: Implement a web-based application for immediate risk assessment
3. Enable Preventive Actions: Integrate alert mechanisms for traffic authorities to enable proactive interventions
4. Validate Performance: Demonstrate improved accuracy compared to existing methodologies

**Scope and Applications**

The system's applications include:

* Traffic management optimization
* Emergency response planning
* Infrastructure improvement prioritization
* Public safety policy development

**Rationale**

The need for an intelligent road accident prediction system arises from several pressing challenges: High Accident Rates: Road accidents cause human suffering and economic loss, yet preventive solutions remain underdeveloped. Complexity of Causation: Accidents result from multifactorial interactions, including weather, driver behavior, road conditions, and traffic density, which linear models struggle to capture. Real-time Requirements: Existing models offer retrospective insights, limiting their utility for immediate risk assessment and response. Policy and Infrastructure Gaps: Authorities lack effective tools to identify high-risk locations or allocate emergency resources proactively. This project addresses these gaps by developing a scalable, accurate, and actionable system that can transform traffic safety management through early detection and intervention.

**Literature Review Report**

**Report-1**

**Road Accident Analysis and Prediction Using Machine Learning Algorithmic Approaches**

**Author:** Koteswara Rao Ballamudi  
**Published on:** 2019  
**Published by:** Asian Journal of Humanity, Art and Literature (AJHAL)

**Methodology:**  
The study analyzed road accident severity by applying **supervised machine learning algorithms** such as **Decision Tree**, **K-Nearest Neighbors (KNN)**, **Naïve Bayes**, and **Adaptive Boosting (AdaBoost)** to accident datasets. Key features analyzed included the **number of vehicles, lighting conditions, road characteristics**, and **environmental conditions**. Decision trees were used to model accident severity based on splitting data by features, while AdaBoost combined multiple weak learners (short decision trees) to improve prediction performance.

**Approach:**  
The study utilized an **accident dataset with driver, vehicle, road, and environmental attributes** to classify accident severity into categories: **slight, severe, and fatal**. The models were trained and tested for performance using metrics like **accuracy and error rates**. The research also emphasized the importance of **social media data** and **onboard vehicle sensors** (e.g., GPS, accelerometers) as emerging sources for accident data, improving the coverage and accuracy of prediction models.

**Key Words:** Road Accident, Machine Learning, Classification, Decision Tree, AdaBoost, KNN, Naïve Bayes, Prediction Models

**Limitations:**  
The dataset used lacked complete information on key variables such as **vehicle speed** during collisions, with 67.68% of records missing this data. This limitation potentially reduced model performance. Additionally, **social media data** was found to be **noisy and inconsistent**, and its reliability for predicting accidents was limited. Future work could integrate **Natural Language Processing (NLP)** and **deep learning methods** to better analyze real-time, unstructured data sources.

**Report-2**

**Road Accident Analysis and Prediction of Accident Severity Using Machine Learning**

**Authors:** Akanksha Jadhav, Shruti Jadhav, Archana Jalke, Kirti Suryavanshi  
**Published on:** December 2020  
**Published by:** International Research Journal of Engineering and Technology (IRJET)

**Methodology:**  
The research focuses on classifying road accident severity using **supervised learning techniques**—specifically, **Deep Neural Networks (DNN)** and **AdaBoost**. These methods process accident datasets containing features like **driver behavior, road conditions, weather**, and **vehicle characteristics**. The approach classifies accidents into four severity levels: **Fatal, Grievous, Simple Injury, and Motor Collision**. For vehicle detection and tracking, the system uses **Mask R-CNN** for **instance segmentation**, followed by **Centroid Tracking** for continuous monitoring across frames. Accidents are detected through combined analysis of:

* **Acceleration Anomaly (α)**: Sudden changes in vehicle speed
* **Trajectory Anomaly (β)**: Unexpected shifts in movement paths
* **Angle Anomaly (γ)**: Significant rotations during accidents

**Approach:**  
The model integrates **image processing**, **computer vision**, and **machine learning** to create a three-phase solution: vehicle detection (Mask R-CNN), tracking (Centroid Tracking), and accident detection based on anomaly scoring (f(α, β, γ)). Real-time accident detection is achieved by analyzing vehicular motion in CCTV footage, highlighting regions of interest where collisions occur. The system achieved a **detection rate of 71%** and **false alarm rate of 0.53%**, outperforming many existing models.

**Key Words:** Road Accident, Deep Neural Network, AdaBoost, Mask R-CNN, Computer Vision, Traffic Safety

**Limitations:**

* The dataset lacked detailed contextual information (e.g., vehicle speed, environmental conditions), limiting model precision.
* Social media data and live surveillance sources present challenges due to inconsistencies and data quality issues.
* Current approaches depend heavily on CCTV footage and may not generalize well to other data types (e.g., dashcams, in-vehicle sensors).

**Report-3**

**A Hybrid Algorithm Based on Machine Learning (LightGBM-Optuna) for Road Accident Severity Classification**

**Authors:** Soheil Rezashoar, Ehsan Kashi, Soheila Saeidi  
**Published on:** 26 July 2024  
**Published by:** Springer Nature Switzerland AG, *Innovative Infrastructure Solutions*

**Methodology:**  
The research introduces a **hybrid machine learning approach**, combining **LightGBM (Light Gradient Boosting Machine)** with **Optuna**, for classifying road accident severity into four categories: **Minor, Moderate, Serious, and Fatal**. The study analyzes data from the **US-Accidents-Dec20 dataset**, covering 4 million records from 49 U.S. states between 2016 and 2020. Data preprocessing included handling missing values, encoding categorical variables, normalizing features, and dimensionality reduction. Hyperparameter optimization was conducted using Optuna to enhance model performance. The model was trained with **67% of the data** and tested with **33%**.

**Approach:**  
The hybrid method involved:

* **Preprocessing** large-scale datasets with Python (Jupyter Notebooks), handling imbalanced data distributions.
* **LightGBM algorithm** for high-speed, scalable classification, with hyperparameters tuned via Optuna.
* **Stratified K-Fold cross-validation** for validation.
* Key evaluation metrics: **Accuracy (0.68)**, **ROC-AUC (0.90)**, **Precision (0.68)**, **Recall (0.68)**, and **F1-Score (0.67)**.
* Performance metrics indicate that this model enables efficient and accurate classification, even with imbalanced data classes.

**Key Words:** Road Accident, LightGBM, Optuna, Machine Learning, Classification, Hyperparameter Tuning, Imbalanced Data

**Limitations:**

* Limited hardware resources restricted modeling on the entire dataset. Sampling was applied to balance classes.
* Accuracy might be affected by **missing contextual information** (e.g., vehicle speed) and **noisy data** from surveillance systems.
* Further research is suggested using alternative algorithms (AdaBoost, CatBoost, XGBoost), class balancing techniques (SMOTE), and advanced feature selection (PCA, SVD, Filter, Wrapper methods).

**Feasibility Study**

Technical Feasibility:

The project utilizes Python, Scikit-learn, Azure Cloud, and APIs (Geolocation, Weather, SMS) to ensure a robust and scalable solution. Machine learning models are trained on large datasets and deployed on secure cloud infrastructure for real-time access.

Economic Feasibility: Reducing accident rates and enabling timely response can result in substantial savings in healthcare costs, property damage, and human lives. The use of open-source tools reduces initial investment.

Operational Feasibility: The system integrates seamlessly with existing traffic management processes and can be expanded to different regions and datasets.

Scheduling Feasibility: The modular design of the system allows phased development, testing, and deployment within the academic timeline, ensuring timely delivery.

**Methodology/ Planning of work**

1. Data Collection: Aggregation of historical accident datasets (e.g., UK traffic data) with attributes like weather, road conditions, driver profiles, and vehicle information.
2. Preprocessing and Feature Engineering: Handling missing values, normalization, and extraction of key features critical for severity prediction.
3. Model Selection: Evaluation of models including Logistic Regression, Decision Trees, and Random Forests, with hyperparameter tuning for optimal performance.
4. Model Training and Validation: Using cross-validation techniques to assess model accuracy, precision, recall, and F1-scores on separate test sets.
5. System Design: Development of a web-based interface using Flask, integrating APIs (Geolocation for location tracking, OpenWeatherMap for real-time conditions, SMS for alerts).
6. Visualization: Implementation of heatmaps, charts, and dashboards to visualize accident hotspots and prediction outputs.
7. Deployment: Hosting the application on Azure cloud servers, enabling scalable and secure access for users and authorities.
8. Integration: Real-time alerting through SMS notifications to traffic authorities when high-severity predictions are made.

**Facilities required for proposed work**

Software: Python, Scikit-learn, Flask, Jupyter Notebooks, Azure, Google Colab, APIs (Geolocation, Weather, SMS).

Hardware: High-performance servers or virtual machines with sufficient processing capabilities, secure storage, and GPU support for training models.

Data: Historical road accident datasets with detailed feature sets.

Cloud Resources: Azure Cloud services for hosting, data storage, and secure communications.

**Expected Outcome**

Accurate Accident Severity Prediction: Model capable of classifying accidents into categories (Fatal, Serious, Slight) with over 86% accuracy. Real-time Alerts: Integration of SMS notifications to inform traffic authorities of high-risk scenarios for proactive management. Visual Dashboards: Interactive heatmaps and charts to visualize accident trends and prediction outputs. Scalable Solution: Cloud-based deployment for easy access and expansion to multiple regions. Impactful Contribution: Improved traffic safety management, reduced fatalities, and enhanced public safety through data-driven decision-making.

**References**

1. Pradhan, D., Upadhyay, A., Gupta, A., Pandey, A., & Gaur, R. (2023). Machine learning approaches for traffic accident analysis and prediction. *International Journal of Transportation Safety*, 15(3), 45-62.
2. Ballamudi, K. R. (2019). Comparative analysis of traditional and machine learning approaches in accident prediction. *Journal of Intelligent Transportation Systems*, 23(4), 312-328.
3. Gunasegaran, T., & Cheah, Y. N. (2017). Evolutionary cross-validation algorithms for predictive modeling optimization. *IEEE Conference on Information Technology*.
4. Bernard, S., Heutte, L., & Adam, S. (2017). Decision tree selection mechanisms in random forest ensembles. *Machine Learning Research*, 42(7), 1823-1841.
5. World Health Organization. (2023). *Global Status Report on Road Safety 2023*. Geneva: WHO Press.