Volume 10, Issue 5, May – 2025 International Journal of Innovative Science and Research Technology
ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25may1803

# Road Accident Prevention System using Machine Learning

Prof. P. Saranya[1]; Gayathri R[2]; Nandhitha A[3]; Keerthika G[4]

[1]Assistant Professor, Department of CSE, Government College of Engineering, Srirangam, Tamilnadu, India,
[2,3,4]UG Student, Department of CSE, Government College of Engineering, Srirangam, Tamilnadu, India

Publication Date: 2025/06/06

**Abstract: Road accidents are a major public safety concern, often resulting in injuries, fatalities, and significant economic loss. Estimating the seriousness of an accident can aid emergency responders and authorities in taking quicker action and enhancing traffic safety. Machine learning offers powerful tools to analyze accident data and make accurate predictions based on various factors such as vehicle type, road conditions, driver behavior, and more. This project uses machine learning to predict how s ever a road accident severity using ensemble classification techniques. The dataset is first preprocessed by handling missing values and encoding categorical variables using Label Encoder. To address class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) is applied, ensuring equal representation of severity classes. The resampled data is then split into training and testing sets. AdaBoost with Random Forest combines the boosting power of AdaBoost with the strong prediction ability of Random Forest to improve classification accuracy. This approach helps in making better predictions even when the original data is imbalanced. Each model's performance is evaluated based on accuracy and the results are compared to identify the most effective model. This model achieved an accuracy of 91.19%, showing its effectiveness in handling imbalanced data and predicting accident severity. The web interface was coded using HTML and CSS with the Flask framework being utilized to connect the trained ML models to the webpage.**

***Keywords:*** *Accident Prevention, Machine Learning, Random Forest, Ada Boost, Severity Prediction.*

**How To Cite**: Prof. P. Saranya; Gayathri R; Nandhitha A; Keerthika G (2025) Road Accident Prevention System using Machine Learning. *International Journal of Innovative Science and Research Technology*, 10(5), 3616-3623. https://doi.org/10.38124/ijisrt/25may1803

## I. INTRODUCTION

Road accidents cause a large number of fatalities, serious injuries, and financial burdens globally, making them a serious public safety concern. By creating a prediction model that automatically categorizes the kind and degree of injuries sustained in different traffic accidents, patterns associated with dangerous incidents could be identified.

Traditional methods for predicting the severity of a traffic collision rely heavily on statistical analysis and rule-based systems [4]. These methods employ historical accident data to uncover patterns and connections among variables such as time of day, road type, vehicle speed, and meteorological conditions. Logistic regression and decision trees are often used to categorize accident severity into minor, major, and deadly. As a result, traditional methods produce broad predictions that may not reflect the individual conditions of an incident, resulting in a delayed or inefficient emergency response and diminished accident prevention effectiveness.

Many models are based on historical accident data, which may not fully represent rapid changes in traffic behavior, road conditions, or weather trends. Moreover, algorithms find it challenging to accurately measure and predict human behaviors like distraction, fatigue, or abrupt choices. The complexity and diversity of real-world traffic situations also make it difficult to create models that are applicable across multiple geographies and scenarios. Furthermore, deploying predictive systems necessitates significant infrastructure, investment, and collaboration across government agencies, which can be a barrier in resource-constrained environments. These constraints underscore the importance of continual progress in data collecting, model development, and policy support in order to increase the accuracy of road accident prediction systems.

The system, when implemented, plans to forecast the severity of road accidents with machine learning models [6] such as Gradient Boost Classiifiers, Decision Tree, Random Forest, SVM, AdaBoost, Cat Boost, AdaBoost with Random Forest and AdaBoost with SVM. These are fitted on the Road Accident dataset to assess the primary attributes like location (longitude and latitude), time, driver statistics (sex and age band), etc. Random forest and AdaBoost are both ensemble

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25may1803

techniques celebrated for their accuracy and stability. Coupled together, these algorithms offer a comparative and complete method of predicting accident severity, contributing to proactively promoting road safety.

A web-based interface was created to render the machine learning (ML) model for predicting road accident severity user-friendly and accessible. The interface was coded using HTML for layout and CSS for presentation, with the Flask framework (a light-weight Python web framework) being utilized to connect the trained ML models to the webpage.

This project aims to study past accident data to find patterns and trends over time, therefore facilitating a better knowledge of the factors influencing road incidents. Predictive models will be created to estimate accident events in particular geographic areas by means of machine learning methods. The aim is to proactively lower accidents by finding high-risk areas and applying focused preventive actions. Weather conditions, traffic flow, and road infrastructure among other things will be taken into account to improve model accuracy. Moreover, the initiative intends to increase general road safety by means of educated planning and strategic resource distribution. A real-time alert system will also be created to inform drivers and appropriate authorities about possible hazards in accident-prone areas, therefore enabling them to reduce the likelihood of collisions and enhance road safety.

## II. RELATED WORK

Numerous studies [1] demonstrate how successful data-driven strategies are in improving road safety. The necessity of automated systems to predict the severity of accidents in Bangladesh was highlighted by Akanksha Jadhav et al. proposed by identified important accident trends using machine learning. Decision trees were used by to examine collisions on the N5 highway.

This study [2] leverages supervised learning techniques, particularly AdaBoost and Deep Neural Networks (DNN), to classify accidents into four severity categories: fatal, grievous, simple injury, and motor collision. Past works, including those by, Sahil Dabhade have explored similar machine learning paradigms like decision trees and feature extraction models for accident analysis.

Ensemble models with high predicted accuracy and resilience, like XG Boost, have drawn interest. XG Boost, a scalable tree boosting technique Xiayoan shen first presented by, has since gained widespread use in research on transportation safety. When compared to alternative approaches, Shanshan wei found that XG Boost significantly improved performance metrics when used to model the severity of injuries in truck crashes. Additionally, XG Boost's usage of feature importance analysis makes it easier to evaluate model outputs, which aids in policy and infrastructure design decision-making [3].

In this research [6], four ensemble learning models based on boosting techniques were employed to predict the severity of injuries resulting from road traffic accidents. The models were built using Python libraries, including NG Boost, Cat Boost, Light GBM, and AdaBoost, along with support from the Scikit-learn framework. The dataset, comprising accident records from National Highway N-5, was sourced from the National Highway and Motorway Police. For model training and evaluation, the dataset was split into three parts: 20% was set aside as a test set for evaluating final model performance, and the remaining 80% was further divided into training and validation sets. The training set was used to train the models, while the validation set assisted in evaluating model performance during hyperparameter tuning. The optimal hyperparameters for each boosting-based model were determined using Bayesian Optimization.

Gaurav prajapathi and avinash provided a comprehensive dataset for additional research and highlighted the impact of outside variables, such as lighting conditions, on accident severity [9]. suggested machine learning strategies to lower road deaths and emphasized the urgent need for predictive systems in emerging nations. Together, this research show how well machine learning predicts accidents and highlight how it may direct preventative safety measures.

The study emphasizes the importance of using predictive analytics to improve road safety strategies. The study [5] compares multiple algorithms such as decision trees, support vector machines, and random forests on datasets of real- world traffic accidents. Their findings suggest that certain algorithms outperform others in accurately classifying accident severity levels, which can significantly aid in timely emergency response and policy-making. This work contributes to the growing body of literature emphasizing the application of advanced computational techniques to mitigate traffic-related hazards.

In the system [7] outlined, a web-based interface was created to render the machine learning (ML) model for predicting road accident severity user-friendly and accessible. The interface was coded using HTML for layout and CSS for presentation, with the Flask framework (a light-weight Python web framework) being utilized to connect the trained ML models to the webpage. After training models such as AdaBoost and more, the top-performing models were stored utilizing the Pickle module in the form of .pkl files. The serialized model files were then loaded into the Flask application in order to facilitate real-time predictions.User interface has the functionality to accept accident-related characteristics like weather, road, driver information, and accident time. The Flask application accepts this input, calls the ML model to make predictions, and returns the predicted severity level of the accident (e.g., slight, serious, or fatal injury). This arrangement enables global access of the model and has the ability to support interactive and effective user experience so that authorities or analysts can make prompt decisions based on predicted severity.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25may1803

Shakil Ahamad and Sayan Kumar Ray explore the prediction of road accidents along with the identification of contributing factors by leveraging explainable machine learning models. The study [9] utilizes various explainable algorithms to analyze traffic and environmental data, providing insights into the key risk factors responsible for road accidents. This approach not only improves prediction accuracy but also enhances transparency, which is crucial for developing effective traffic management strategies and safety policies. Their work represents a significant advancement in applying machine learning techniques that balance predictive performance with explainability in the domain of road safety.

## III. PROPOSED SYSTEM

Figure 1 focuses on a road accident prevention system using machine learning. The process begins with data collection from various sources, including **t**raffic conditions**,** Speed limit**,** longitude, latitude, lighting conditions**,** and driver- related data**.** This raw data is then sent through a data preprocessing stage to clean and structure it for use. The cleaned data is divided into training and testing datasets**,** which are used to train a proposed machine learning (ML) model specifically-AdaBoost with Random Forest. Once trained, the model performs severity prediction based on user input regarding current road and driver conditions. The system then classifies the predicted accident severity as either major or minor**.** This system enables proactive measures to minimize road accidents by using intelligent predictions based on historical data. encoding to categorical features and label encoding to the target variable. Handling Class Imbalance: Uses SMOTE (Synthetic Minority Over-sampling Technique) to balance the classes in the training data. The dataset is split into two parts: 70% is used to train the model, while the remaining 30% is reserved for evaluating its performance.

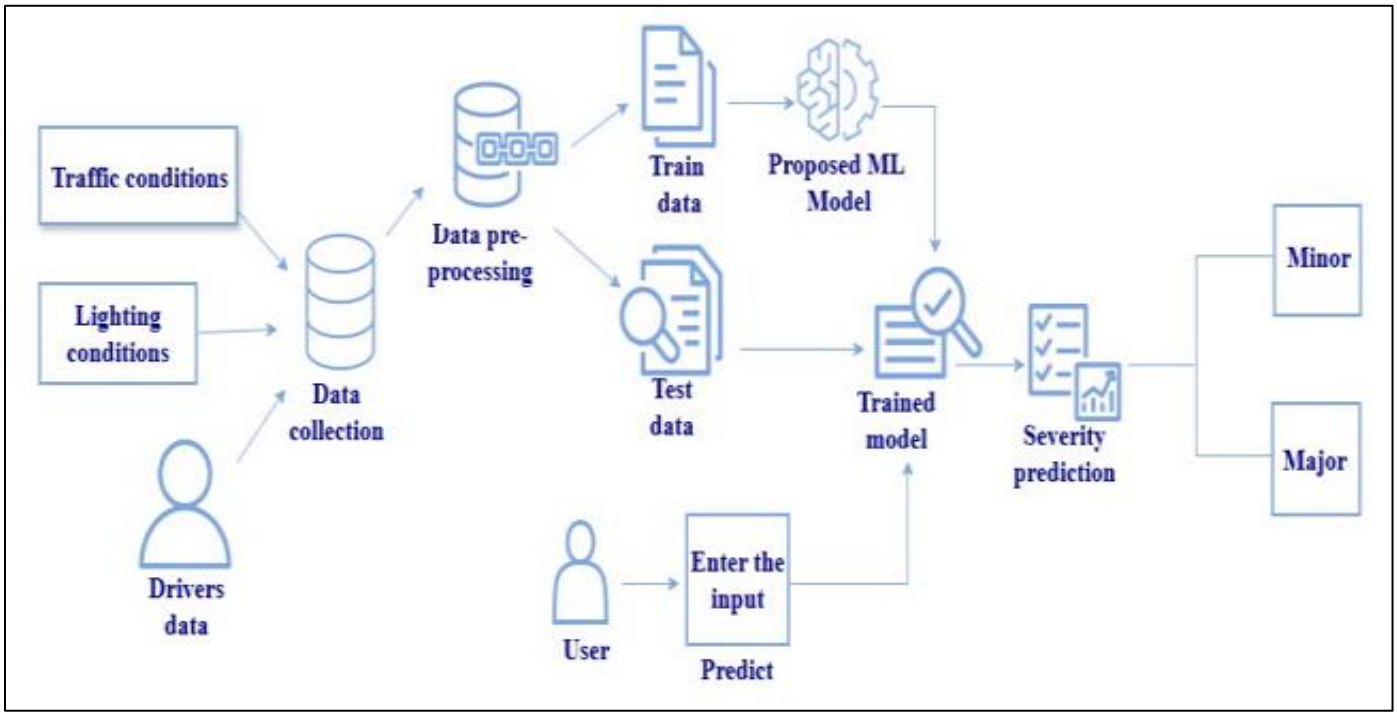


Fig 1 System Architecture

## IV. MODULES

➢ *Data Collection*

Dataset collection is the process of gathering and organizing data that will be used for analysis, modeling, or research. For analysis, the Kaggle Road Accident dataset was imported using Python's panda's module. It provides crucial parameters such as longitude, latitude, driver sex, time, driver age range, driver behavior and accident severity. These characteristics aid in understanding accident trends and predicting road safety results.

- Reads a dataset from a CSV file: RoadAccident.csv

➢ *Data Pre-Processing:*

It means cleaning, transforming, and organizing raw data into a usable format so that a machine learning model can understand and learn from it properly. Handling missing values: Identifies and handles missing values, replacing an and Unknown with Nan, and imputes missing values either by the mode or a placeholder. Drops irrelevant or low value columns such as Time. Label Adjustment :Simplifies the target class Accident_ Severity by merging Fatal injury into Slight injury for binary or reduced-class classification. Feature Selection and Encoding: Select important features like Days of week, Sex of driver, Age band of driver, etc. Applies one-hot encoding to categorical features and label encoding to the target variable. Handling Class Imbalance: Uses SMOTE (Synthetic Minority Over-sampling Technique) to balance the classes in the training data. The dataset is split into two parts: 70% is used to train the model, while the remaining 30% is reserved for evaluating its performance.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25may1803

➢ *Model Training and Evaluation:*

The model training and testing process involves evaluating four different machine learning algorithms to classify the severity of road accidents. During training, each classifier is applied to the training data to learn the underlying patterns and connections between the input features and the target variable. Ada Boost was implemented using Random Forest as the base estimator to enhance classification performance. Once e trained, the models are tested on the unseen test set to evaluate their generalization ability. Predictions are made using the. predict () method, this model's performance is evaluated using accuracy as the metric, these accuracy scores are then stored for comparison, offering a clear comparison to identify the top-performing model on the dataset.

➢ *Web app Integration:*

A web-based interface was created to render the machine learning (ML) model for predicting road accident severity user-friendly and accessible. The interface was coded using HTML for layout and CSS for presentation, with the Flask framework being utilized to connect the trained ML models to the webpage. After training models AdaBoost with RF, the top-performing models were stored utilizing the Pickle module in the form of .pkl files. The serialized model files were then loaded into the Flask application.

➢ *Evaluation Metrics:*

Various evaluation metrics are vital for evaluating the performance of statistical, machine learning, and deep learning models. These indicators are critical in measuring the effectiveness of a suggested model in a study. Evaluation metrics including as accuracy, precision, recall, and the F1 - score are important for determining a model's prediction or classification efficacy.

➢ Accuracy

Accuracy, expressed as a percentage, signifies the proportion of images that are accurately predicted among all predictions made. Equation (1) defines accuracy as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

The accuracy score compares the model's total number of correct predictions (TP + TN) against its total number of predictions. The initials TP, TN, FP, and FN represent "truepositive," "true negative," "false positive," and "false negative" accordingly.

➢ *Precision*

The precision, indicating the proportion of truly positive outcomes among the predicted positive instances, is calculated using this equation (2):

$$Precision = \frac{TP}{TP + FP}$$

➢ *Recall*

Recall measures how well positive cases are identified, calculated as the ratio of true positives to the sum of true positives and false negative, as shown in Equation (3).

➢ *F1-Score*

The study typically use metrics like the F1-score, which is calculated by taking the harmonic mean of the model's precision and recall, to assess the performance of their models. The formula for this equation will be [4]:

$$F_1\ Score = 2 \times \frac{Recall \times Precision}{Recall\ \times Precision}$$

➢ *Implementation:*

The Road Accident Severity Prediction System was implemented using Python 3.10, supported by essential libraries such as pandas, numpy, matplotlib, scikit-learn, cat boost, and imblearn. The development was carried out on a system with Windows 10, Intel Core i5 processor, and 4GB RAM, using Jupyter Notebook for data processing and Visual Studio Code for backend and frontend integration.

The process began with data preprocessing of the accident dataset (Road.csv), which includes features such as vehicle type, driver gender, number of passengers, road surface type, speed limit, day of the week, and light conditions. The dataset was cleaned by removing missing values and encoding categorical variables using Label Encoder. To handle class imbalance—where some accident severity levels occurred less frequently—SMOTE (Synthetic Minority Over- sampling Technique) was used. This technique generated synthetic samples for the minority classes, ensuring balanced representation of all severity levels during training.

The features (X) and labels (y) were separated and the data was split into training and testing sets with a 70:30 ratio. For modeling, AdaBoostClassifier with RandomForestClassifier as its base estimator was used. This ensemble approach combines the boosting strength of AdaBoost with the predictive power of Random Forest, enhancing accuracy and robustness even with initially imbalanced data.

The model was trained and evaluated using performance metrics such as Accuracy, Precision, Recall, and F1-Score. Among these, accuracy was used as the primary metric for comparison, and the final model achieved a high accuracy of 91.19%, demonstrating its capability to make reliable predictions on unseen data.Figure 2 depicts the recall is the highest, which means the model rarely misses major accidents The precision correctly identifies most major accidents, helping focus attention where it matters most. F1-score being high confirms the model has a good balance between being sensitive and precise.Accuracy above 91% reflects strong overall performance on the dataset.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25may1803

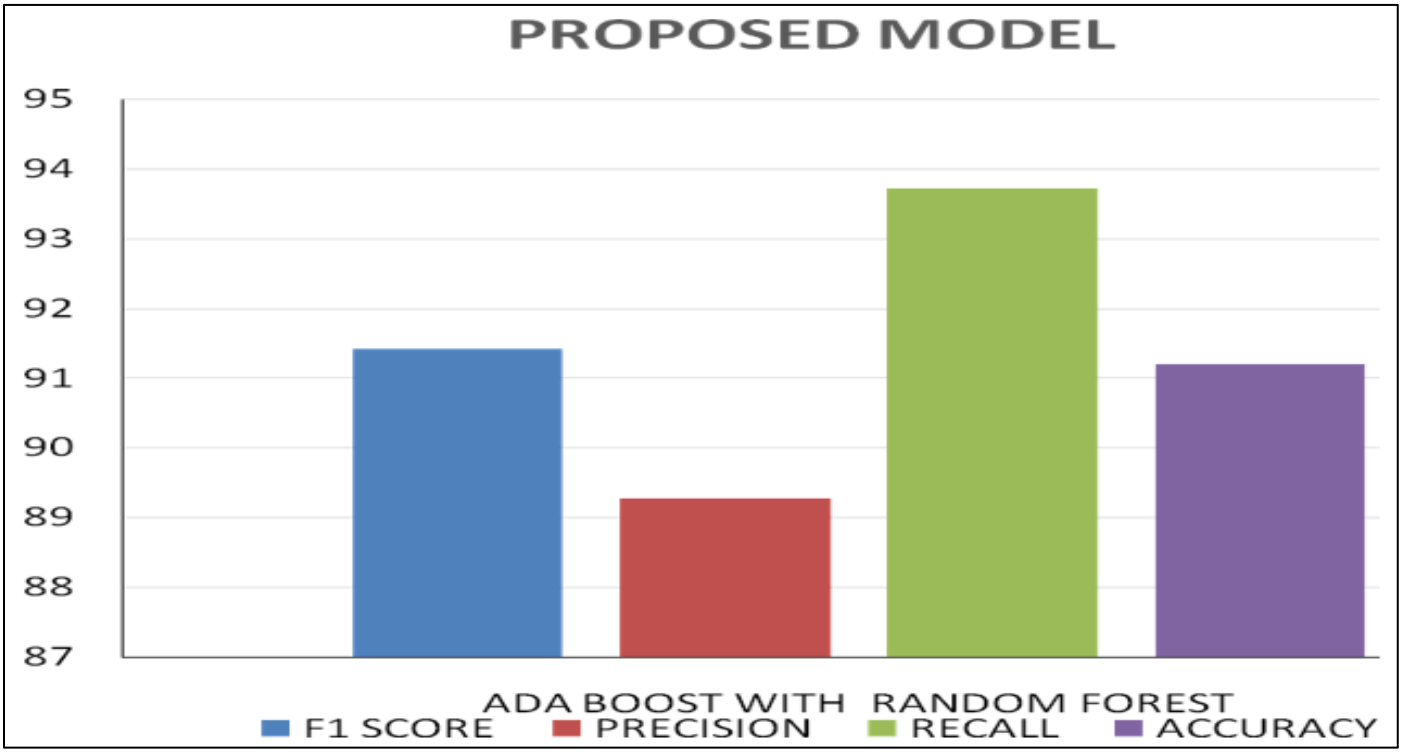


Fig 2 Performance Metrics

Figure 3 depicts the confusion matrix is based on a classification model using AdaBoost with Random Forest as the base estimator.

- **True Positive (TP)** = 674 → Major accidents correctly predicted as Major.
- **True Negative (TN)** = 688 → Minor accidents correctly predicted as Minor.
- **False Positive (FP)** = 89 → Minor accidents incorrectly predicted as Major.
- **False Negative (FN)** = 48 → Major accidents incorrectly predicted as Minor.

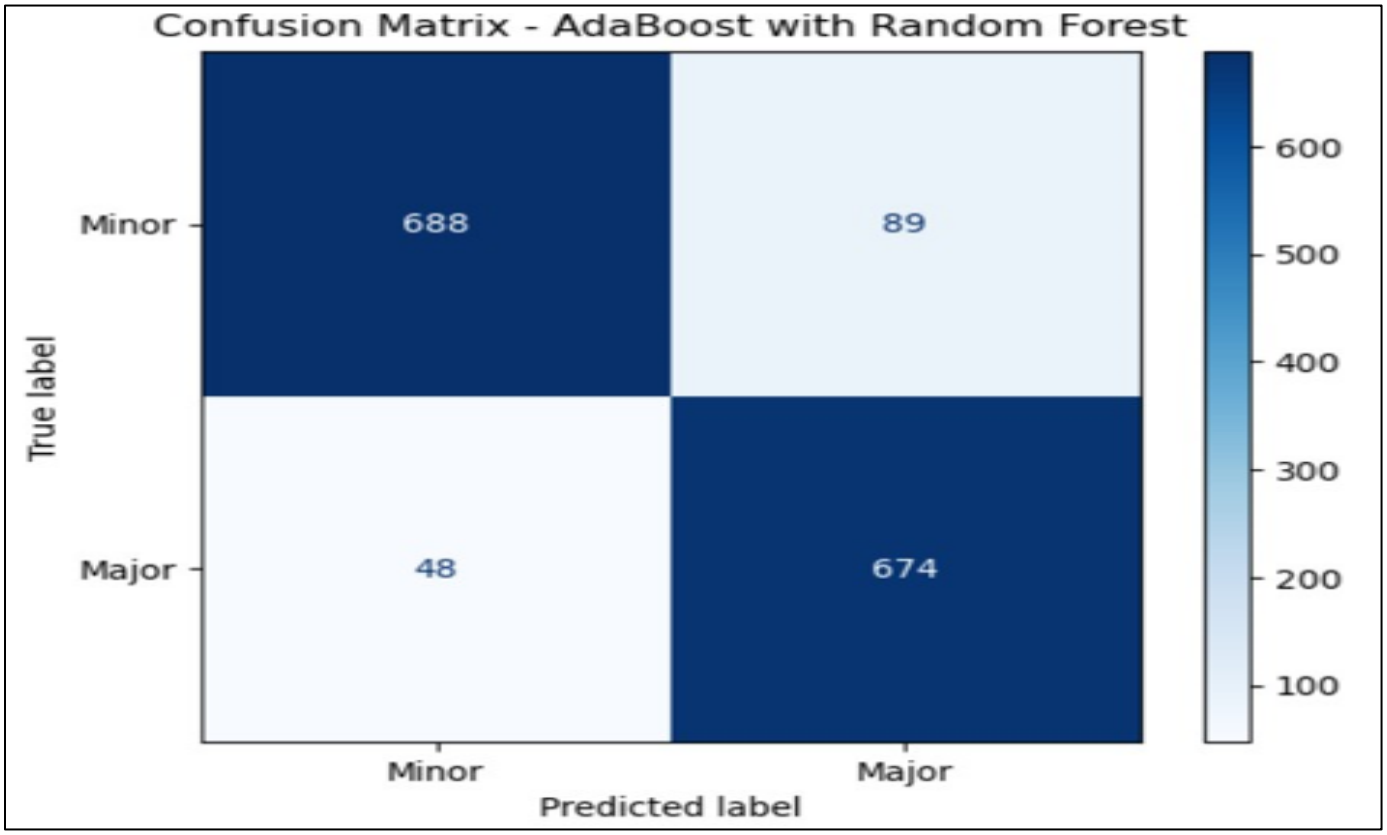


Fig 3 Confusion Matrix

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25may1803

For deployment, Figure 4 depicts the trained model is integrated into a Flask web framework and connected with a frontend interface developed using HTML, CSS, and JavaScript. The frontend allows users to enter accident-related inputs such as latitude, longitude, driver sex, vehicle type, number of passengers, road type, speed limit, light conditions, and day of the week. Upon submission, these inputs are sent to the backend where the model processes the data and returns the predicted severity of the accident [8].

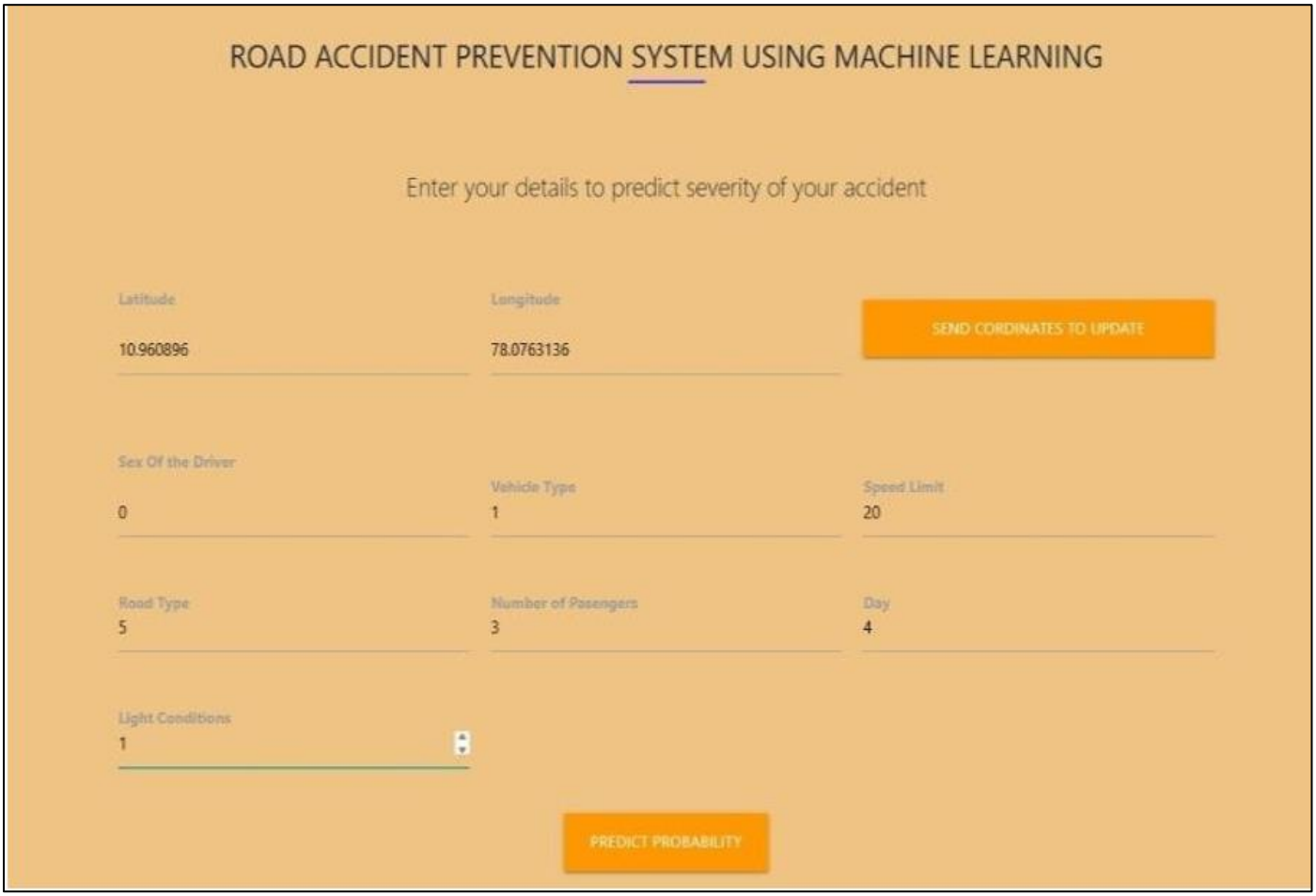


Fig 4 Road Accident Severity Prediction

The intuitive and accessible design enables users like road safety officers, drivers, and policy makers to assess the potential severity of road accidents and make informed, proactive decisions to improve road safety.

## V. RESULTS AND DISCUSSION

Figure 5 presents a comparative analysis of six machine learning models—SVM, KNN, Decision Tree, Random Forest, Ada Boost, and a Proposed Model(Adaboost with RF)—based on four performance metrics: F1 Score, Precision, Recall, and Accuracy.Among these models, the Proposed Model demonstrates the most balanced and superior performance, especially in terms of Recall and Accuracy.Overall, the chart highlights the effectiveness of the Proposed Model in achieving both high accuracy and reliability.

Table 1 Comparison of Model Performance Metrics

| MODEL | ACCURACY | PRECISION | RECALL | F1-SCORE |
|---|---|---|---|---|
| Svm | 52.97 | 53.18 | 49.13 | 51.08 |
| Knn | 75.92 | 99.74 | 51.94 | 68.31 |
| Decision Tree | 82.12 | 83.36 | 80.24 | 81.77 |
| Random Forest | 90.79 | 89.02 | 93.06 | 90.99 |
| Adaboost | 62.91 | 58.88 | 85.45 | 69.72 |
| Proposed Model | 91.19 | 89.27 | 93.72 | 91.41 |

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25may1803

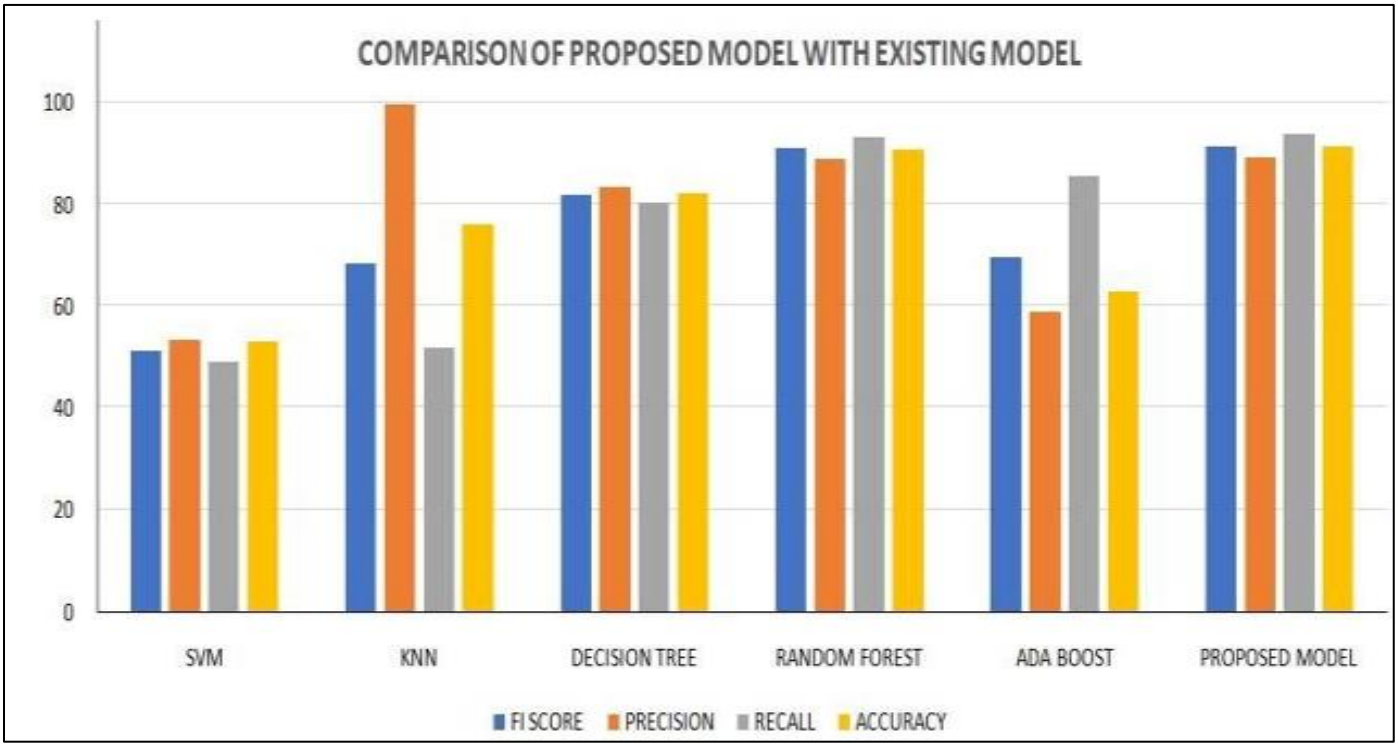


Fig 5 Model Comparison

## VI. CONCLUSION AND FUTURE WORK

An hybrid-based machine learning approach was implemented to develop a road accident prediction system, aiming to classify and predict accident injury severity effectively. The system utilized powerful machine learning algorithms—Ada Boost with Random Forest—to evaluate and compare their performance on a historical accident dataset.

The experimental results revealed that AdaBoost with Random Forest outperformed the other models**,** achieving the highest accuracy. This indicates its superior ability to capture complex patterns in the data by combining the strengths of boosting and the robustness of Random Forest. The final model classified accident severity into two groups: major and minor. The use of machine learning techniques—particularly AdaBoost with Random Forest—proves to be an effective solution for improving the accuracy and reliability of road accident severity prediction systems and reducing the impact of road traffic incidents.

For future upgrades to the Accident Severity Prediction System, the current interface and performance analysis charts provide a solid foundation. The current method can be enhanced by including real-time data sources such as GPS, weather updates, and traffic congestion feeds to produce dynamic and location-aware predictions. In addition, the user-friendly interface might be improved with smartphone compatibility and voice input to make it more accessible in emergency situations. Another potential improvement would be to include accident heatmaps and alert systems to warn drivers in high-risk areas.

## REFFERENCES

[1]. Akanksha Jadhav, Shruti Jadhav, Archana jalke," Road accident analysis and prediction of accident severity using machine learning." International Research Journal of Engineering and Technology (IRJET), ISSN-2395-(0056- 0072),2020.

[2]. Sahil Dabhade, Sai mahale, Avinach Chitkala, "Road accident analysis and prediction using machine learning." International Journal for Research in Applied Science & Engineering Technology, IC:45.98, ISSN:2321-9653,2020.

[3]. Shanshan wei, Xiayoan shen," Application of XG Boost for hazardous material road transport accident severity analysis", IEEE pages:206806 – 206819,2020.

[4]. Arun Venkat, Gokulnath M, Guru Vijey K.P, Irish Susan Thomas, "Machine learning based analysis for road accident prediction." International Journal of Emerging Technology and Innovative Engineering, 6(2),2020.

[5]. Shakil Ahamad,Sayan Kumar Ray,Md Akbar Hossain,"A Comparitive study of machine learning algorithms to predict road accident severity",International Conference on Ubiquitous computing and communications,2021.

[6]. Sheng Dong, Arshad Hossain, Irfan Ullah," Predicting and analyzing road traffic injury severity using boosting based ensemble learning models", International journal of environmental research and public health, 19(5), 2925,2022.

[7]. Koteswararao Kodepogu , Vijaya Bharathi Manjeti "Machine Learning for Road Accident Severity Prediction" Mechatronics and Intelligent Transportation Systems, pp. 211–226, 2023.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25may1803

[8]. Gaurav Prajapati, Avinash, Lav Kumar, Smt. Rekha S Patil,"Road Accident Prediction Using Machine Learning." Journal Of Scientific Research & Technology, 48 -59,2023 ISSN 2583-8660,2023.
[9]. Shakil Ahamad,Sayan Kumar Ray,"A Study on road accident prediction and contributing factors using explainable machine learning models" Transportation Research interdisciplinary perspectives,19,100814,2023.