Improving Road Safety Through Machine Learning Based Severity Prediction

Akil Chakravarthi A
Computer Science and Engineering
Easwari Engineering College Chennai,India akilannakodi@gmail.com

Kamalesh M
Computer Science and Engineering
Easwari Engineering College Chennai,India kamalesh052004@gmail.com

Kanishka Devi S
Computer Science and Engineering Easwari Engineering College Chennai,India
kanishkadevi2004@gmail.com

Dr.S. Kayalvizhi
Computer Science and Engineering Easwari Engineering College Chennai,India
kayalvizhi.s@eec.srmrmp.edu.in

***Abstract*—An efficient predictive system that can precisely categorize accident severity and allow for targeted interventions is necessary given the increased frequency of traffic accidents. This study divides traffic accidents into two groups: ”Serious Injury” and ”Slight Injury”. It uses advanced machine learning methods to do this.. The prediction model was developed using a big dataset that included environmental parameters, road conditions, driver demographics, and vehicle attributes. A range of machine learning techniques, including ensemble and non- ensemble models, were investigated in order to determine the most accurate approach. The XGBoost classifier was determined to be the top-performing model following hyperparameter tuning optimization, attaining a high prediction accuracy of 99road surface features are important in influencing the severity of accidents, according to the model's feature importance analysis. The outcomes show how effective machine learning frameworks are at promoting preventive safety measures and enhancing the distribution of resources for traffic control. This prediction model has the potential to greatly increase road safety by offering real- time, data-driven insights. It can also help traffic authorities and legislators make well- informed decisions to lessen the effect of serious traffic incidents.**

***Index Terms*—Traffic Accident Severity, Machine Learning, XGBoost, Predictive Modeling, Road Safety, Accident Prediction.**

I. INTRODUCTION

Road traffic accidents remain one of the leading causes of fatalities and injuries worldwide, presenting a significant danger to public health and the economy. According to estimates from the World Health Organization (WHO), 1.3 million people lose their lives in road accidents each year, and millions more have non-fatal injuries that frequently result in permanent disability. These alarming data underscore the need of identifying and mitigating variables that contribute to the

Identify applicable funding agency here. If none, delete this.

severity of road accidents. Developing strategies to minimise casualties, reduce economic loss, and improve overall road safety requires an understanding of the contributing factors to accident severity. To gain meaningful insights, however, sophisticated data-driven methodologies must be used because accident severity prediction based on multiple influencing factors is intrinsically complex.

Recent advances in machine learning(ML) have made it feasible to utilize historical data on road accidents to forecast accident severity with a high degree of accuracy. ML is especially well-suited to modelling the multifactorial nature of traffic accidents because of its capacity to unearth intricate relationships within data. A number of interrelated factors influence how serious an accident is: driver demographics, vehicle characteristics, road attributes, and environmental con- ditions. Since ML models can handle these complex relation- ships between multiple variables, they are an effective tool for analysing the severity of traffic accidents. Additionally, the automation of predictions made possible by ML-based techniques allows for timely responses that may save lives.

This study uses a dataset that includes a variety of features, such as the type of vehicle involved, weather, road surface characteristics, driver age and experience, and accident cir- cumstances, to investigate the application of machine learning (ML) algorithms to forecast the severity of road accidents. These characteristics offer a thorough understanding of the elements influencing the result of an accident. The study's objective is to evaluate the effectiveness of various machine learning techniques,

including Support Vector Machines, Ran- dom Forest, and Gradient boosting SVM Machines—perform in predicting the target variable, which is accident severity, which is divided into two categories: "Slight Injury" and

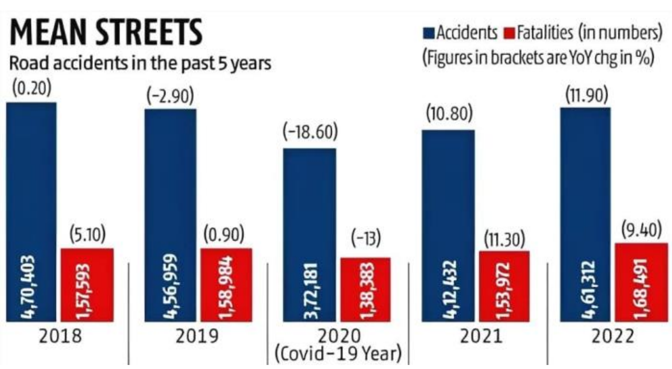


Fig. 1. Summary of Road Accidents in India

"Serious Injury." This classification is extremely important because it forms the basis for the allocation of resources and the prioritisation of emergency medical services, which may lessen the degree of injuries incurred. An essential part of the entire process is data preparation, which makes sure the dataset is ready for machine learning modelling. If left unchecked, missing data, inconsistent results, and superfluous features can significantly affect the model's performance. Systematic data preprocessing techniques include imputation of missing data, normalization of numerical features, and encoding of cate- gorical variables were employed. Apart from improving the model's understanding of the data, this meticulous preparation helped to reduce problems like bias and overfitting that are fre- quently seen in predictive modelling assignments utilising real- world datasets. Another crucial component of this research is feature selection, which establishes the model's predictive ability. The performance of the model might be negatively impacted by an extensive feature set since it can add noise and computational inefficiencies. To choose a subset of features that have the biggest influence on accident severity prediction, Recursive feature elimination (RFE) and correlation analysis were among the techniques employed. This method improves the model's interpretability while simultaneously increasing its computing efficiency. The best machine learning model for the given classification task was found by training and assessing a variety of models. The benchmark model was the XGBoost classifier, which is renowned for its resilience and capacity to handle both continuous and categorical variables. The ensemble nature of the model, which involves training multiple decision trees and aggregating the results, ensures resilience against overfitting and yields higher accuracy, partic- ularly in complex datasets. Additionally, models like Support Vector Machines and Gradient Boosting were employed for comparison analysis. Hyperparameter tuning was carried out by combining grid search and random search approaches to make sure that each model operates at its optimal settings for this specific scenario. The model's performance was evaluated using metrics such as accuracy, precision, recall, and F1- score, with particular attention paid to the area under the receiver operating characteristic curve (AUC-ROC). Collec- tively, these indicators provide insight into the model's abilityto differentiate between the two severity classifications. Given the imbalanced nature of real-world traffic data, where minor injuries are more prevalent than major ones, additional metrics, such as precision-recall curves, were also utilized to assess the model's efficacy. This research sought to interpret the factors associated with higher accident severity in addition to devel- oping a predictive model. Factors like ”Weather conditions,” ”Road surface conditions,” and ”Type of vehicle” are among those that are most likely to affect the severity of an acci- dent, as indicated by the feature importance scores obtained from models such as XGBoost. For the purpose of creating interventions aimed at lowering serious injuries, legislators and traffic safety authorities need to have such insights. Reiterating the significance of vehicle safety checks, for example, or putting better road maintenance strategies into practice can all help reduce the chance of serious accidents. As can raising driver awareness during certain weather conditions. In addition to using machine learning models to forecast accident severity, this work is unique since it thoroughly assesses many approaches to identify the best fit for this particular use case. Through methodical assessment of several models and metic- ulous parameter tuning, the research offers a well-informed foundation for choosing the model that performs the best. Even though XGBoost had the best accuracy, each model's interpretability and computational efficiency were taken into account to provide a comprehensive assessment. This increases our understanding of the different ways that machine learning methods might be used to enhance road safety in real world situations. It is anticipated that this project will have a major impact on the real world, especially in the areas of emergency response planning and road safety management. By enabling efficient resource dispatching based on incident severity, the capacity to forecast accident severity can significantly improve the effectiveness of emergency services. Furthermore, traffic management authorities can proactively implement policies and measures to lower the incidence of serious injuries by identifying important contributing factors. Therefore, this work has the potential to enhance not only the immediate response to traffic accidents but also the long-term strategies meant to improve road safety in general.

II. RELATED WORK

[1] The project focuses on predicting the severity of traffic accidents using text data from police-filed First Information Reports (FIRs). The challenge was designed as a text classi- fication task by the researchers using a bi-directional Long Short-Term Memory (LSTM) model. obtains a perfect F1 score of 90build an extensive database of traffic accidents. de- pends on the consistency and quality of FIR data, which might differ by location. [2] Using a six-year dataset from Southeast Michigan, this study uses machine learning models such as Decision Tree (DT) and Random Forest (RF) to predict crash severity.. It also identifies key crash severity variables. DT offers computational efficiency. Decision Tree and Random Forest outperform other classifiers in terms of accuracy. Data imbalance issues (through aggressive sampling) affected the

overall model accuracy .[3] This study analyzes traffic accident severity using the Support Vector Machine (SVM) model incorporating multiple variables like age, gender, weather, and lighting conditions. SVM achieved 83.7Considers diverse factors for a holistic understanding of accident severity. May struggle with scalability or handling larger datasets. [4] The study focuses on accident severity prediction in California using XGBoost with data balancing techniques like random undersampling and SMOTENC oversampling. Effectively han- dles unbalanced data. SHAP interpretability method offers insights into influential accident factors. The use of complex balancing techniques could increase the computational burden.

[5] Using data from traffic accidents in Leeds, this study evaluates Random Forest, Support Vector Machine (SVM), and Artificial Neural Network (ANN) models to forecast the severity of traffic accidents. The best accuracy was 93Random Forest. Provides detailed performance comparisons between models. Model accuracy and applicability may be region- specific. [6] An Artificial Neural Network (ANN) model is applied to predict traffic accident severity using various factors such as vehicle type, weather conditions, and time of the accident. High model accuracy (80range of crash- related features. ANN models may not scale

efficiently with larger datasets or require extensive computational resources. [7] This study integrates K-Nearest Neighbors (KNN) and DBSCAN algorithms to improve traffic accident severity prediction ac- curacy. Hybrid model improves overall prediction accuracy. Successfully detects clusters and outliers enriching accident severity analysis. Complex ensemble models are resource-intensive. [8] This research analyzes highway accident severity using gradient boosting, highlighting the factors that influ- ence crash outcomes. Gradient Boosting classifier outperforms other models. SHAP analysis helps identify influential factors contributing to crash severity. Limited generalizability outside the dataset used (Thailand). [9] This study uses a Graph Neural Network (GNN) to predict crash severity, capturing relationships among crash records that conventional models may overlook. GNN outperforms other models like XGBoost and ANN in accuracy. Identifies hidden patterns in the data. High computational cost compared to traditional models.

[10] This study proposes a Deep Hybrid Attention Network (DHAN) to predict crash severity and enhance emergency response management. DHAN outperforms baseline models with an AUC of 0.820. Captures both spatial and temporal variations for dynamic prediction. Requires significant data processing and computational resources. [11] In order to fore- cast collision severity, this study used a two-phase ensemble model, addressing data labeling and imbalance difficulties while concentrating on distracted driving. efficiently uses SMOTE to handle unlabeled and unbalanced data. reaches a remarkable 99.6Combines hotspot analysis with machine learning to improve road safety. Kernel Density Estimation (KDE) effectively identifies dangerous zones. Results are specific to the UK and may not apply to other regions.

III. EXISTING SYSTEM

Conventional statistical approaches used to forecast the severity of traffic accidents, such as logistic regression, lin- ear discriminant analysis, and generalized linear models, are the foundations of the current methodology. The simplicity, interpretability, and ease of implementation of these tech- niques have led to their widespread use. They are capable of modelling the relationships between accident severity and contributing factors with effectiveness, giving insights into the characteristics that have a major influence on results. Policy- makers and traffic management authorities can easily access these techniques because they facilitate simple data analysis. These traditional methods' interpretability is one of their main benefits. Making direct connections between the predictor variables and the severity outcome provides decision- makers with important information. Furthermore, conventional models are typically easier to implement with common statistical tools and require less processing power. These methods do, however, have a number of important drawbacks. The biggest flaw is that they are unable to accurately represent intricate, non-linear relationships between variables, which are frequently present in datasets pertaining to traffic accidents. Such oversimplified modeling can lead to inaccurate predictions, particularly when there is significant interaction between multiple factors. Fur- thermore, especially in high-dimensional datasets, traditional models are more prone to problems like multicollinearity and overfitting. They also lack the flexibility needed to automati- cally discover intricate patterns in large datasets, which limits their predictive power compared to modern machine learning techniques. These techniques might not be able to precisely forecast the seriousness of accidents in dynamic real-world situations because of the intricate relationships between several variables.

IV. PROPOSED WORK

By employing machine learning (ML) techniques, the algorithm in question estimates the severity of traffic accidents and classifies them as either ”Slight Injury” or ”Serious Injury.” The researchers used a bi-directional Long Short-Term Memory (LSTM) model to build the task as a text classification task among the several variables that impact the severity of an accident in order to overcome the shortcomings of conventional statistical models. By taking into account variables including road characteristics, environmental circumstances, driver demographics, vehicle type, and accident causes, the method offers a more complete picture of outcome prediction.

A range of machine learning models, including Support Vector Machines, Gradient Boosting, and XGBoost, are used in the construction of this system; each model is evaluated for accuracy and overall performance. Finding the model that predicts accident severity the best is the aim; this model will then be suggested for use in real-world scenarios. The system is optimised for precision, dependability, and generalisability thanks to this comparative method. The suggested system begins with pre-processing the dataset to guarantee the best possible quality of data. This covers managing missing values, encoding category variables, and normalizing numerical characteristics. To narrow down the set of input variables to the ones that are most important for precise prediction, feature selection techniques are applied. This guarantees computational efficiency while improving model performance and interpretability.

Along with precise forecasts, the selected machine learning model offers insights into the variables that most strongly affect the severity of accidents. To identify the major contributors, such as the type of vehicle, road surface, and weather, feature importance analysis is carried out. With the help of these insights, authorities are better equipped to decide whether to target driver education programs, improve infrastructure, or implement extra safety measures when conditions are particularly risky.

The suggested system also seeks to enhance the distribution of emergency resources by offering accurate and timely forecasts. The capacity to precisely assess an accident's severity in an emergency can aid in allocating dispatch resources according to priority, potentially saving lives. MoreoverBy analyzing past data to find patterns and trends, the system can assist policymakers in putting preventive measures into action.

All things considered, it is projected that the proposed machine learning (ML)-based system would greatly improve the capacity to forecast the intensity of traffic accidents, assisting with more focused interventions, better road safety decision-making, and more efficient emergency response preparation. It is an effective tool that can improve safety, lessen the overall effect of traffic accidents, and ultimately save lives.

V. SYSTEM ARCHITECURE

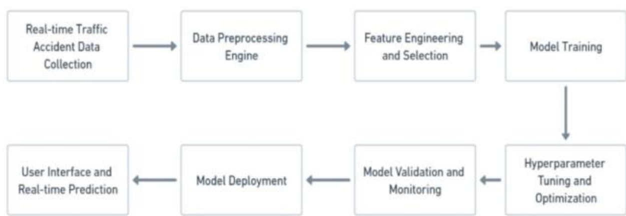


Fig. 2. System Architecture Diagram

A. Data Collection and Storage

Information about traffic accidents is gathered from a variety of sources, such as incident logs, government databases, and road safety publications. Features like vehicle types, road surface characteristics, environmental factors, and driverdemographics are among the features included in the data collection. Scalability and data integrity are guaranteed by using a secure database management system (DBMS) to store this data effectively. Data updates, retrieval, and integration into the machine learning pipeline are all made possible by the database. Enforcing proper encryption and access control mechanisms prioritises data quality and privacy.

B. Data Processing Pipeline

The pipeline for data processing is an essential part of the architecture of the system; it transforms unprocessed data into a format that can be used for modelling and analysis. To ensure standardisation, data transformation is necessary, followed by feature engineering to extract more pertinent variables and data cleaning to eliminate errors and inconsistencies. One-hot or label encoding is used to encode categorical variables, and normalisation is applied to numerical variables to place them into comparable ranges. Once the dataset has been processed, it is separated into test, validation, and training sets so that models can be trained and assessed.

C. Machine Learning Framework

The machine learning framework, which trains several mod- els to predict the severity of traffic accidents, is the main component of the system. A range of techniques, including Support Vector Machines, Gradient Boosting, and XGBoost, are used in early testing. Using Python-based machine learning libraries like Scikit-learn, the models are developed in an integrated development environment (IDE) such as Jupyter Notebook. Once the optimal model has been selected based on performance criteria, the model is prepared for deployment and trained using the whole dataset. To identify the most important factors in predicting the severity of an accident, the method employs a feature significance analysis.Model Deployment A crucial stage in allowing the trained model to make predictions in practical applications is model deployment. A web-based API is used to deploy the chosen model, making it easy for users to access and integrate with the user interface. The portability and effective operation of the model in various environments are guaranteed by a containerisation strategy based on Docker. In addition, deployment entails keeping an eye on the model’s output, regularly updating it with fresh information, and guaranteeing that it can reliably produce predictions in real time.

D. User Interface and Accident Visualization

The users can input specific accident-related details and receive a prediction regarding the accident's severity thanks to an interactive user interface (UI) that offers a custom input prediction feature. Furthermore integrated into the user interface is a map-based visualisation module that lists all of the accidents that have happened in each state of India. In order to help authorities identify high-risk areas and put preventive measures in place, the map offers a clear and understandable view of the accident distribution. Better decision-making and

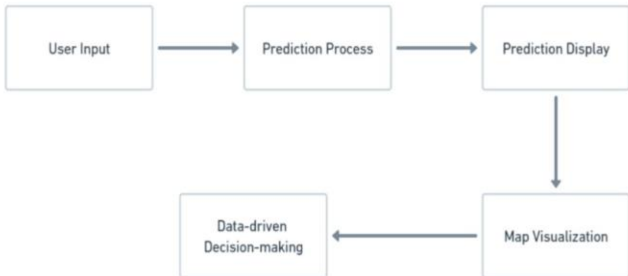


Fig. 3. Accident Visualization Process

C. Model Selection and Training accident prevention tech- niques are made possible by the combination of predictive modelling and data visualization.

VI. METHODOLOGY

A methodical approach is utilized for data processing, feature engineering, model training, and evaluation in order to precisely forecast the seriousness of traffic accidents. From the initial data preparation to the model evaluation, this section describes the steps used to create a reliable machine learning model, guaranteeing peak performance and reliability.

A. Data Preprocessing Techniques

To ensure that the dataset is suitable for machine learning, preparation is a crucial step. In the process's data cleaning stage, missing values are handled using imputation tech- niques. To make categorical features compatible with machine learning models, such as "Type of vehicle" and "Weather conditions," they are encoded using label encoding or one- hot encoding. Normalisation is a useful technique for stan- dardising numerical features and bringing them into a similar range, which improves model convergence during training. To enhance data quality and lower noise, redundant records are also eliminated, and format inconsistencies are fixed.

B. Feature Selection and Dimensionality Reduction

Only the most relevant characteristics for predicting acci- dent severity are retained by feature selection, which reduces the likelihood of overfitting and improves the interpretability of the model. To find relationships between features and find multicollinearity, a correlation matrix is first used. To rank the features according to how important they are for predicting the target variable, Recursive Feature Elimination (RFE) is then used. This makes it easier to identify a subset of crucial characteristics, such "Weather conditions," "Type of vehicle," and "Driver age band," that most effectively contribute to the model's performance. Reducing the number of dimensions in a model maximises its computational efficiency by ensuring that only the essential variables are used.

C. Model Selection and Training

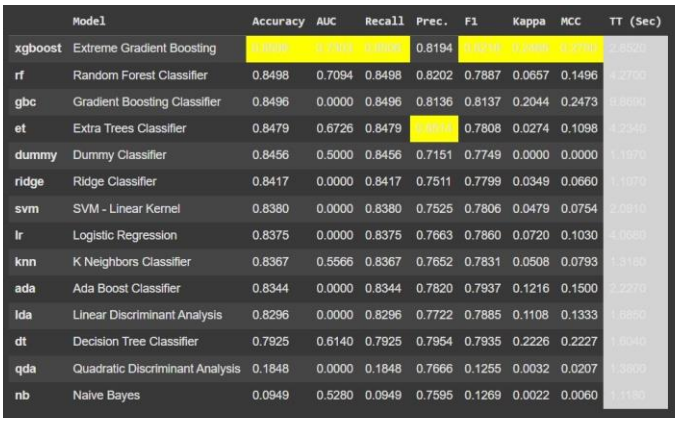
The optimal match for forecasting accident severity is found by comparing many machine learning techniques. Training models that use the processed dataset include XGBoost, Gra- dient Boosting, and Support Vector Machines (SVM). All models undergo an initial training with default settings, and the process is executed in Python utilising Scikit- learn. The most appropriate model for the classification task is then determined by comparing the models' performances. Given its robustness in managing intricate, non-linear interactions, XGBoost, an ensemble learning technique, is chosen for additional improvement.

D. Hyperparameter Tuning and Cross-Validation

Optimizing model performance requires careful adjustment of the hyperparameters. The optimal set of hyperparameters for every model is discovered using methods like grid search and random search. To improve accuracy, variables like the maximum depth, learning rate, and number of trees in XGBoost, Gradient Boosting, and others are changed. Specifically, k-fold cross-validation is performed to ensure that the model does not overfit and that it generalizes well to new data. This facilitates the evaluation of model consistency across dataset subsets.

E. Model Evaluation Metrics

The model's performance is assessed using a variety of assessment indicators. The main metrics used to assess how successfully the model distinguishes between "Slight Injury" and "Serious Injury" are accuracy, precision, recall, and F1 score. The Area Under the Receiver Operating Characteristic Curve (AUC-ROC), which accounts for the imbalance in class distribution, is also used to assess the model's effectiveness in class distinguishing. Confusion matrices are also used to drive future model improvement and offer insights into categorisation errors. By using many metrics, the model's predictive power is thoroughly assessed, improving the model's dependability in practical settings.



	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
xgboost	Extreme Gradient Boosting	[illegible]	[illegible]	[illegible]	0.8194	[illegible]	[illegible]	[illegible]	2.8520
rf	Random Forest Classifier	0.8498	0.7094	0.8498	0.8202	0.7887	0.0657	0.1496	4.2700
gbc	Gradient Boosting Classifier	0.8496	0.0000	0.8496	0.8136	0.8137	0.2044	0.2473	[illegible]
et	Extra Trees Classifier	0.8479	0.6726	0.8479	[illegible]	0.7808	0.0274	0.1098	4.2340
dummy	Dummy Classifier	0.8456	0.5000	0.8456	0.7151	0.7749	0.0000	0.0000	1.1970
ridge	Ridge Classifier	0.8417	0.0000	0.8417	0.7511	0.7799	0.0349	0.0660	1.1070
svm	SVM - Linear Kernel	0.8380	0.0000	0.8380	0.7525	0.7806	0.0479	0.0754	2.0910
lr	Logistic Regression	0.8375	0.0000	0.8375	0.7663	0.7860	0.0720	0.1030	[illegible]
knn	K Neighbors Classifier	0.8367	0.5566	0.8367	0.7652	0.7831	0.0508	0.0793	1.3160
ada	Ada Boost Classifier	0.8344	0.0000	0.8344	0.7820	0.7937	0.1216	0.1500	2.2270
lda	Linear Discriminant Analysis	0.8296	0.0000	0.8296	0.7722	0.7885	0.1108	0.1333	[illegible]
dt	Decision Tree Classifier	0.7925	0.6140	0.7925	0.7954	0.7935	0.2226	0.2227	1.6040
qda	Quadratic Discriminant Analysis	0.1848	0.0000	0.1848	0.7666	0.1255	0.0032	0.0207	[illegible]
nb	Naive Bayes	0.0949	0.5280	0.0949	0.7595	0.1269	0.0022	0.0060	[illegible]

Fig. 4. Before Tuning

ALGORITHM

The XGBoost classifier, a powerful and incredibly efficient machine learning tool used for both classification and regres- sion issues, is the method employed for this project. A quick and effective way to create gradient-boosted decision trees is with XGBoost. It works by producing trees one after the other and correcting errors made by the previous trees. XGBoost produces a more accurate and dependable model by concen- trating on difficult-to-predict data points through this iterative boosting method. XGBoost is a great option for applications with large-scale, high-dimensional data, such as cybersecurity, where it is crucial to detect malicious activity from logs, network traffic, and other sources. It is very excellent at managing sparse data and big datasets. XGBoost uses methods like regularization, downsizing, and handling missing data to increase prediction accuracy. XGBoost produces a strong prediction model by adding new trees in order to minimize the residuals from earlier trees.

A. Cost Function Splitting Criteria

Log Loss: When it comes to binary classification tasks, XGBoost optimizes the log loss function. It computes the true class's negative log-likelihood and is defined as:

$$Log\ Loss = \frac{1}{n}\sum_{i=1}^{n}[yi\ log(pi) + (1 - yi)\ log(1 - pi)]$$

• Gain (Information Gain): Each node of a tree in XGBoost uses a splitting criterion that is based on optimizing the information gain, which quantifies the decrease in uncertainty following the split. The best feature and splitting threshold are chosen by calculating information gain using the loss function.

B. Evaluation Metrics

Accuracy: The percentage of real results—true positives and true negatives—among all the cases that were looked at.

$$Accuracy = \frac{Number\ of\ Correctly\ Identified\ Cases}{Total\ Number\ of\ Instances}$$

Precision: The percentage of all positive projections that were accurately predicted.

$$Precision = \frac{True\,Positives}{True\,Positives + False\,Positives}$$

Recall: The ratio of true positive forecasts to all actual positives.

$$Recall = \frac{True\,Positives}{True\,Positives + False\,Negatives}$$

F1 Score: The precision and recall harmonic mean, which offers a balance between the two.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

	precision	recall	f1-score	support
0	0.98	1.00	0.99	2080
1	0.85	0.86	0.86	2079
2	0.86	0.84	0.85	2090
accuracy			0.90	6249
macro avg	0.90	0.90	0.90	6249
weighted avg	0.90	0.90	0.90	6249

Fig. 5. After Tuning

VII. RESULT ANALYSIS

The XGBoost classifier performed exceptionally well in pre- dicting the severity of traffic accidents, attaining an astounding 99the dataset, the model's high degree of accuracy shows that it can efficiently distinguish between the ”Slight Injury” and ”Serious Injury” categories. By means of hyperparameter tweaking and cross-validation, XGBoost was refined to effec- tively manage the intricate and non-linear interactions present in the data, hence surpassing other machine learning models assessed for the project.

The model performed quite well in terms of other evaluation criteria in addition to accuracy, with a high precision and recall balance shown by an F1 score that was very near to 1. This suggests that in addition to correctly classifying the severity of accidents, the model also reduced false positives and false negatives. Important factors that significantly contribute to accident severity were identified by the feature importance analysis, including weather, vehicle characteristics, and type of road surface. The XGBoost model is well-suited for practical uses in resource allocation and traffic accident prevention due to its robust prediction powers and capacity to manage big datasets.

VIII. CONCLUSION

The study successfully illustrates the application of machine learning techniques to predict the severity of traffic accidents,

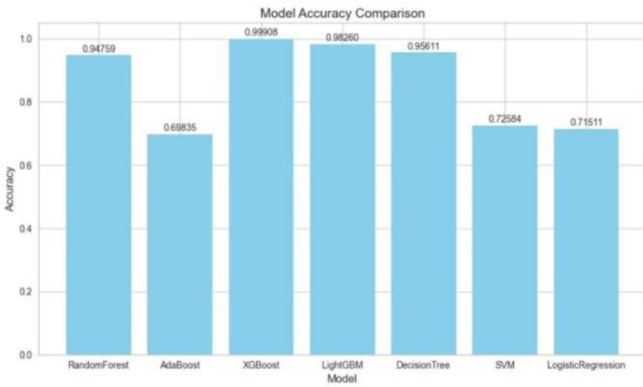


Fig. 6. Comparison of Model Results

offering a more effective alternative to traditional statistical models. The suggested model uses a range of data, such as driver demographics, vehicle specifications, environmental conditions, and road qualities, to reflect the complex relation- ships that influence accident outcomes. Following a compari- son of other algorithms, XGBoost was shown to be the most accurate model, offering significant insights into the variables affecting the severity of accidents. Variables including weather, vehicle type, and kind of road surface have a considerable impact on the severity of accidents, according to the feature importance study. These insights can assist decision-makers in putting specific policies into place that will lessen the frequency and seriousness of traffic accidents. Additionally, the proposed system offers authorities a workable way to ef- ficiently allocate resources and prioritise emergency response. It includes an interactive user interface and visualisation tools. This work has important implications for improving road safety and reducing accident- related casualties, in addition to improving the predictive power of traffic accident analysis. Subsequent studies may concentrate on incorporating real-time data and investigating sophisticated deep learning models to enhance forecast precision and flexibility in ever changing.

IX. REFERENCES

[1] References K. K. Ahirwar, O. Mishra and G. Ramadurai, ”Determining Road Crash Severity from Police First Informa- tion Reports,” 2022 14th International Conference on Com- munication Systems NETworkS (COMSNETS), Bangalore, India, 2022, pp. 10.1109/COMSNETS53615.2022.9668585.
[2] 854-858, doi: S. Hossain and D. Valles, ”Traffic Safety through Machine Learning: A Study of Crash Severity Fac- tors,” 2024 IEEE World AI IoT Congress (AIIoT), Seat- tle, WA, USA, 2024, pp. 0016-0023, doi: 10.1109/AI-IoT61789.2024.10579015. [3] S. M. Sadeque Hamdan, W.
K. Awan and K. A. Ghuzlan, ”Enhancing Traffic Acci- dent Severity Prediction Using Artificial Intelligence Tech- niques,” 2023 24th International Arab Conference on In- formation Technology (ACIT), Ajman, United Arab Emi- rates, 2023, 10.1109/ACIT58888.2023.10453808. [4] pp. 1-6, doi: J. Lu, Z. Huang and L. Yang, ”Road Traffic Ac- cident Severity Prediction under Unbalanced Data,” 2023 4th International Conference on Computer, Big Data and Artificial Intelligence (ICCBD+AI), Guiyang, China, 2023,
pp. 650-654, doi: 10.1109/ICCBD-AI62252.2023.00119. [5]
I. E. Mallahi, A. Dlia, J. Riffi, M. A. Mahraz and H. Tairi, ”Prediction of Traffic Accidents using Random Forest Model,” 2022 International Conference on Intelligent Sys- tems and Computer Vision (ISCV), Fez, Morocco, 2022,
pp. 1-7, doi: 10.1109/ISCV54655.2022.9806099. [6] S. M.
S. Hamdan, S. Barakat, K. H. Mahfouz and K. A. Ghu- zlan, ”Traffic Accident Severity Prediction Model using AI,” 2023 Advances in Science and Engineering Technology In- ternational Conferences (ASET), Dubai, United Arab Emi- rates, 2023, pp. 10.1109/ASET56582.2023.10180861. [7] 1-
5, doi: B. V. S. Kiran, M. A. Kowshik, A. L. Manohar,
D. R. Siddardha, M. Kavitha and S. Kavitha, ”Enhanc- ing Accident Prediction Through Integrated KNN and DB- SCAN Algorithms for Superior Accuracy,” 2024 8th In- ternational Conference on Inventive Systems and Con- trol (ICISC), Coimbatore, India, 2024, pp. 423-428, doi: 10.1109/ICISC62624.2024.00078. [8] B. Vanishkorn and W.
Supanich, ”Crash Severity Classification Prediction and Fac- tors Affecting Analysis of Highway Accidents,” 2022 9th

International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA), Tokoname, Japan, 2022, pp. 1-6, doi: 10.1109/ICAICTA56449.2022.9932998. [9] K.

A. Sattar, I. Ishak, L. S. Affendey and S. N. B. Mohd Rum, ”Road Crash Injury Severity Prediction Using a Graph Neural Network Framework,” in IEEE Access, vol. 12, pp. 37540- 37556, 2024, doi: 10.1109/ACCESS.2024.3373885. [10] M. T.

Kashifi, ”Deep Hybrid Attention Framework for Road Crash Emergency Response Management,” in IEEE Transactions on Intelligent Transportation Systems, vol. 25, no. 8, pp. 8807-

8818, 10.1109/TITS.2024.3376653. Aug. 2024, doi: [11] S.

Bag, S. Maity and S. Sarkar, ”Crash severity analysis in distracted driving using unlabeled and imbalanced data: A novel approach using Robust Two-Phase Ensemble Predictor,” 2022 International Conference on Data Analytics for Business and Industry (ICDABI), Sakhir, Bahrain, 2022, pp. 88-92, doi: 10.1109/ICDABI56818.2022.10041646. [12] U. Gupta, V. MK

and G. Srinivasa, ”A Comprehensive Study of Road Traffic Accidents: Hotspot Analysis and Severity Prediction Using Machine Learning,” 2022 IEEE Bombay Section Signature Conference (IBSSC), Mumbai, India, 2022, pp. 1-6, doi: 10.1109/IBSSC56953.2022.10037449. [13] S. S. Reddy, Y. L.

Chao, L. P. Kotikalapudi and E. Ceesay, ”Accident analysis and severity prediction of road accidents in United States using machine learning algorithms,” 2022 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), Toronto, ON, Canada, 2022, pp. 1-7, doi: 10.1109/IEMTRON- ICS55184.2022.9795852. [14] Y. Lu and R. Kuang, ”A Driver

Injury Prediction Model based on Genetic Algorithm and BP Neural Network,” 2023 7th International Conference on Transportation Information and Safety (ICTIS), Xi’an, China, 2023,

pp. 1984-1989, doi: 10.1109/ICTIS60134.2023.10243825.